

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

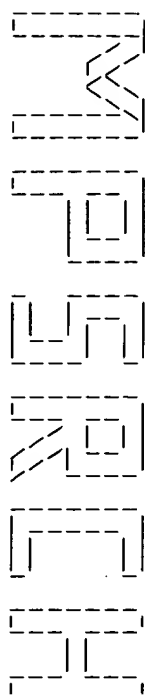
Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparc_ntp n.a. n.a. Smith-Waterman search, using a protein database
which has been backtranslated into n.a. using IUPAC symbols

Run on: Fri Apr 16 14:29:02 1999; Maspar time 95.81 Seconds

Tabular output not generated. 944,391 Million cell updates/sec

Title: >US-09-011-797-1
Description: (1-932) from US09011797.seq
Perfect Score: 4660
N.A. Sequence: 1 CTCTCCAGCGTGTTCAGCAG.....AAAAAAAAAAGAAATTC 932
Comp: GAGAGGCGCAGCAAGTCGTC.....TTTTTTTTTTCCTTAAG

Scoring table: TABLE bkttranslated2
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 131922 seqs, 48541980 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

a-geneeq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 57.111; Variance 273.971; scale 0.208

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	199	4.3	17 24	W25163	Human opioid receptor	4.50e+03
2	196	4.2	17 25	W25163	Rat orphanin FQ recep	7.17e+03
3	196	4.2	17 24	W25162	Human opioid receptor	7.17e+03
4	181	3.9	17 25	W25220	Rat orphanin FQ recep	7.06e+02
5	175	3.8	1719 16	R92100	Human RIZ allele D283	1.73e+01
6	170	3.6	562 14	R70491	Leucocytotoxin protoso	3.61e+01
7	170	3.6	712 25	W30749	Rat vrs21 gene produc	3.61e+01
8	166	3.6	4987 3	R10834	Rianodin receptor.	6.48e+01
9	161	3.5	392 27	W30091	FIN13 serine/threonin	1.33e+00
10	160	3.4	58 4	PA0312	PIG brain thalamus po	1.54e+00
11	160	3.4	249 17	R85900	Human pp32 protein.	1.54e+00
12	157	3.4	412 19	W03626	Human thyrotropin GR	2.36e+00
13	155	3.3	182 17	R85902	Mouse pp32 partial se	3.13e+00
14	155	3.3	182 4	R21740	pp35-related protein.	3.13e+00

c	15	155	3.3	707 16	R79912	Human nucleolin.	3.13e+00
	16	156	3.3	2986 8	R44432	eryA region polypepti	2.72e+00
	17	151	3.2	542 28	W36490	Human TUP1 protein.	5.48e+00
	18	148	3.2	875 28	W34987	Bankia goudii endoglu	8.31e+00
	19	146	3.1	118 2	R12349	Toxoplasma gondii F2	1.09e+01
	20	143	3.1	162 6	R30641	pCENP-B-1 protein jac	1.65e+01
	21	145	3.1	289 8	R52633	Guinea pig pR-30, 30	1.25e+01
	22	143	3.1	594 7	R34936	CENP-B.	1.65e+01
	23	146	3.1	725 27	W39165	Human RHAMM protein.	1.09e+01
	24	144	3.1	729 1	R05457	KEX1 amino acid sequ	1.44e+01
	25	141	3.0	19 24	W00941	Acidmax acidc extens	2.16e+01
	26	142	3.0	98 24	W00951	bHx Max with acidic	1.89e+01
	27	140	3.0	226 6	R46485	Rat GAP-43.	2.47e+01
	28	140	3.0	226 6	R25332	Sequence encoded by r	2.47e+01
	29	140	3.0	226 6	R05907	Rat neuronal growth r	2.47e+01
	30	140	3.0	226 6	R30015	Rat GAP-43.	2.47e+01
	31	140	3.0	226 6	R46496	Rat GAP-43.	2.47e+01
	32	140	3.0	227 1	R05908	Mouse neuronal growth	2.47e+01
	33	140	3.0	283 27	W37482	Mouse liver cancer-or	2.47e+01
	34	138	3.0	414 20	W10529	Saccharomyces cerevis	3.23e+01
	35	140	3.0	905 26	W31186	Human p160 polypeptid	2.47e+01
	36	140	3.0	1132 17	R97866	Chicken leucocytozoan	2.47e+01
	37	140	3.0	1135 26	W31185	Human p160 polypeptid	2.47e+01
	38	136	2.9	27 23	W11629	Anti-apoptotic protei	4.21e+01
	39	135	2.9	27 17	R94328	Anti-thermogenic pep	4.80e+01
	40	135	2.9	252 1	P94369	Fusion protein congl.	4.80e+01
	41	135	2.9	288 29	W41574	Human cardiac tropoin	4.80e+01
	42	135	2.9	352 23	W22437	Frog neurogenic diffe	4.80e+01
	43	135	2.9	360 19	W03627	Human follicle stiml	4.80e+01
	44	135	2.9	484 29	W43910	Heterodera glycines s	4.80e+01
	45	135	2.9	1427 2	R10534	Human 160KD mediator	4.80e+01

ALIGNMENTS

RESULT 1
ID W25163 standard: peptide: 17 AA.
AC W25163;
DE 08-DEC-1997 (first entry)
DT Human opioid receptor-like 1 receptor ligand, nociceptin, fragment.
KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
KW memory; attention; sensory perception; learning; homeostasis;
KW hypoaesthesia; nociceptin; endorphin; dynorphin A.
OS Synthetic.
PN W09707208-A1.
PD 27-FEB-1997.
PF 14-AUG-1986; BE0087.
PR 15-AUG-1995; US-002366.
PA (ULBR) UNIV LIBRE BRUXELLES.
PI Meunier J, Mollereau C, Parmentier M, Vassart G;
DR WPI: 97-165292/15.
DR N-PSDB: T79888.
PT Novel ligand for the opioid receptor-like receptor, nociceptin -
PT has pro-nociceptive properties, useful for treating or preventing
PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
PT etc.
PS claim 7; Page 27; 48pp; English.
CC W25162-W25164 are fragments of a ligand to human opioid receptor-like 1
CC (ORL1) receptor, designated nociceptin, which resembles the endorphin
CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
CC sequences and antibodies can be used as new types of drugs in the
CC control of various behaviours or functions. The inhibitors can be used
CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
CC learning and memory, homeostasis, hyperalgesia, hypoaesthesia and/or
CC sensory perception.
SQ Sequence 17 AA;
Query Match 4.3%; Score 199; DB 24; Length 17;
Best Local Similarity 49.0%; Pred. No. 4.50e+03;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 1 f s e f m r q y l v l s m q s s q
 Dt 1 f s e f m r q y l v l s m q s s q
 OY 421 TTGAGGCTTATGAGGACGACGCTGCTGAGCACTGACGACCA 471
 Ct F S E F M R Q Y L V L S M Q S S Q

RESULT 2
 ID W25219 standard; peptide; 17 AA.
 AC W25219;
 DT 12-JAN-1998 (first entry)
 DE Rat orphanin FQ receptor-binding peptide inhibitor.
 KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
 KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
 OS morphine induced analgesia; methadone specific opioid receptor; ss.
 PN W09707/212-A1.
 PD 27-FEB-1997.
 PR 12-AUG-1996; U13305.
 PR 03-NOV-1995; US-553058.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Bunzow JR, Civealli O, Grandy DK, Grisell JE, Mogil JS;
 DR MPM: 97-165296/15.
 PT Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
 treatment of locomotor disease.
 PS Claim 12; Page 48; 68pp; English.
 CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (ORF)
 receptor.
 CC Highly specific peptides that bind the ORF receptor were identified as
 ORF receptor inhibitors. The peptides can be used to antagonise a
 physiological effect of an opioid in an animal. The peptides antagonise
 opiate induced hypothermia and morphine induced analgesia in animals.
 CC They may also be used in the diagnosis and treatment of locomotor
 disease. The peptides may also be used in the design of a methadone
 specific opioid receptor (MSOR) in drug design and for the isolation
 of endogenous receptors for anti-opioid agonists and antagonists found
 in plasma, serum, lymph, cerebrospinal fluid etc.
 CC Sequence 17 AA;

Query Match 4.2%; Score 196; DB 25; Length 17;
 Best Local Similarity 56.9%; Pred. No. 7.17e-03;
 Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dt 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGAGGCTTATGAGGACGACGCTGCTGAGCACTGACGACCA 471
 Ct F S E F M R Q Y L V L S M Q S S Q

RESULT 3
 ID W25162 standard; peptide; 17 AA.
 AC W25162;
 DT 08-DEC-1997 (first entry)
 DE Human opioid receptor-like 1 receptor ligand, nociceptin, fragment.
 KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity; disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 OS Synthetic.
 PN W09707/208-A1.
 PD 27-FEB-1997.
 PR 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368.
 PA (ULBR) UNIV LIBRE BRUXELLES.
 PI Neunier J, Molleireau C, Parmentier M, Vassart G;
 DR N-PDB: T79888.
 PT Novel ligand for the opioid receptor-like receptor, nociceptin -
 has pro-nociceptive properties, useful for treating or preventing
 diseases related to e.g. stress, hyperalgesia, locomotor activity;

PT etc.
 PS Claim 5; Page 27; 48pp; English.
 CC W25162-W25164 are fragments of a ligand to human opioid receptor-like 1
 CC (ORL1) receptor, designated nociceptin, which resembles the endorphin
 CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
 CC sequences and antibodies can be used as new types of drugs in the
 CC control of various behaviours or functions. The inhibitors can be used
 CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
 CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
 CC learning and memory, homeostasis, hyperalgesia, hypoaesthesia and/or
 CC sensory perception.
 CC Sequence 17 AA;

Query Match 4.2%; Score 196; DB 24; Length 17;
 Best Local Similarity 56.9%; Pred. No. 7.17e-03;
 Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dt 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGAGGCTTATGAGGACGACGCTGCTGAGCACTGACGACCA 471
 Ct F S E F M R Q Y L V L S M Q S S Q

RESULT 4
 ID W25220 standard; peptide; 17 AA.
 AC W25220;
 DT 12-JAN-1998 (first entry)
 DE Rat orphanin FQ receptor-binding peptide inhibitor.
 KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
 KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
 OS morphine induced analgesia; methadone specific opioid receptor; ss.
 PN W09707/212-A1.
 PD 27-FEB-1997.
 PR 12-AUG-1996; U13305.
 PR 03-NOV-1995; US-553058.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Bunzow JR, Civealli O, Grandy DK, Grisell JE, Mogil JS;
 DR MPM: 97-165296/15.
 PT Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
 treatment of locomotor disease.
 PS Claim 12; Page 48; 68pp; English.
 CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (ORF)
 receptor.
 CC Highly specific peptides that bind the ORF receptor were identified as
 ORF receptor inhibitors. The peptides can be used to antagonise a
 physiological effect of an opioid in an animal. The peptides antagonise
 opiate induced hypothermia and morphine induced analgesia in animals.
 CC They may also be used in the diagnosis and treatment of locomotor
 disease. The peptides may also be used in the design of a methadone
 specific opioid receptor (MSOR) in drug design and for the isolation
 of endogenous receptors for anti-opioid agonists and antagonists found
 in plasma, serum, lymph, cerebrospinal fluid etc.
 CC Sequence 17 AA;

Query Match 3.9%; Score 181; DB 25; Length 17;
 Best Local Similarity 56.9%; Pred. No. 7.06e-02;
 Matches 29; Conservative 10; Mismatches 12; Indels 0; Gaps 0;
 Db 1 f g g f t g a r k s a r k l a n g
 Dt 1 f g g f t g a r k s a r k l a n g
 OY 364 TTGAGGCTTATGAGGACGACGCTGCTGAGCACTGACGACCA 471
 Ct F S E F M R Q Y L V L S M Q S S Q

RESULT 5
 ID R2100 standard; Protein; 1719 AA.
 AC R2100;
 DT 19-MAY-1996 (first entry)

Query Match	3.68;	Score 170;	DB 25;	Length 712;
Best Local Similarity	42.28;	Pred. No. 3.61e-01;		
Matches	35;	Conservative		

[illegible]

ID R10834 standard; Protein; 4987 AA.
 AC R10834.
 DT 26-APR-1991 (first entry)
 DE Rianodin receptor.
 KW Rianodin receptor gene; calcium release modulator; tranquiliser;
 OS antagonist.
 PN Oryctolagus cuniculus.
 PF J03011098-A.
 PD 18-JAN-1991.
 PE 07-JUN-1989; 144569.
 PR 07-JUN-1989; JP-144569.
 PA (MITU) Mitsubishi Kasei Corp.
 WP1: 91-062003/09.
 DR N-PDB: Q10613.
 PT New Rianodin receptor, genes encoding it and its prep. - useful as
 PT calcium release modulator for tranquilisers and for assaying
 PT calcium antagonists; 18pp; Japanese.
 PS Disclosure; Fig 1; 18pp; Japanese.
 CC RNA contg. poly(A) was prepared from rabbit skeletal muscle
 CC endoplasmic reticulum. From the obtd. poly(A) mRNA, a cDNA bank
 CC corresp. to it was prepared by random primer method, oligo (dT)
 CC primer method, and primer extension method. A cDNA was obtained by
 CC screening with a DNA probe (see Q10614-15).
 CC cDNA into an expression vector. By introducing the obtd.
 CC The product an expression vector, vector PRK7 was formed.
 CC Sarcoplasmic reticulum which triggers constriction of skeletal muscle.
 CC Therefore, the receptor is useful as tranquiliser and assay series for
 CC screening of calcium antagonist.
 SO Sequence 4987 AA;

	Query Match	3.6%	Score 166;	DB 3;	Length 4987;
	Best Local Similarity 43.2%;		Pred. No. 6,486-01;		
	Matches	41;	Conservative	15;	Mismatches 38; Indels 0; Gaps 0;
Dd	1 k r k l g v d g e e e l v p e p e				
Dt	13266 hytnaarnaaaytngngtngdgygngargargarytngtncmgargcgc 13325				
Oy	243 CCTAAGGATGACCCGCGTGTACAGAGTGCTGTCAAGCCGAGACGACGACCTCGAGC 302				
Ot	L K R M P R V R S V Q A R D A E P E A				
Dd	e p e p e p e k a d e				
Dt	13326 ngatcngarcngarcngarcngaraaagcnggygarg 13360				
Oy	303 AGATGCAGAGCTCGTCGCGATGTAGGCCGATGTAGG 337				
Ot	D A E P V A D				

RESULT.

ID	W30091 standard; Protein; 392 AA.
AC	W30091
DT	07-APR-1998 (first entry)
DE	FIN13, serine/threonine phosphatase.
KW	FIN13, serine/threonine phosphatase; hyperproliferative disorder;
KM	fibroblast growth factor-inducible protein; collagen-homology domain;
KM	uncontrolled proliferation; neoplasm; tumour; therapy.
OS	MS sp.
FH	Key
FT	Domain
FT	Location/Qualifiers
FT	1..100
FT	/note="collagen-homology domain"
FT	108..169
FT	/note="acidic-box domain"
FT	174..352
FT	/note="serine/threonine phosphatase catalytic domain"
PD	MO9335018-A1.
PN	25-SEP-1997.
PF	21-MAR-1997; U05075.
PR	21-MAR-1996; US-622339.
PR	21-MAR-1996; US-013792.
PA	(UNYNY) UNIV NEW YORK STATE.
PI	Basiglio C, Guttridge MA;
PI	Wpi: 97.480223/44.
DR	N-P5DB; T91793.
PT	Fibroblast growth factor-inducible protein (FIN13) serine-threonine
PT	phosphatase - used to treat a disease or disorder associated with
PT	uncontrolled proliferation
PS	Claim 2; Page 84-85; 116pp. English.
CC	This sequence represents the protein of the invention, and is a
CC	fibroblast growth factor-inducible protein (FIN13) serine/threonine
CC	phosphatase, which has a collagen-homology domain, an acidic box, a
CC	serine/threonine phosphatase domain and a charged domain characteristic
CC	of a putative nuclear localisation sequence. A host cell transformed with
CC	the DNA encoding this sequence is used for the recombinant production of
CC	FIN13. A vector containing the DNA encoding this sequence can be used to
CC	increase the level of expression of FIN13, especially in an animal.
CC	or disorder associated with uncontrolled proliferation in an animal.
CC	FIN13 can also be administered directly to treat uncontrolled
CC	proliferation, e.g. hyperproliferative disorders, neoplasms and tumours.
CC	A labelled antibody against this protein, or an antisense oligonucleotide
CC	against the DNA encoding this sequence can be used to detect and
CC	quantitate FIN13 in a biological sample.
SO	Sequence 392 AA;

Query Match	3.58;	Score 161;	DB 27;	Length 392;
Best Local	Similarity 39.18;			
Matches	52;	Conservative	Pred. No. 1.33e+00;	

[illegible]

RESULT 10
ID P40312 standard; Protein; 58 AA.
AC P40312;
DT 18-AUG-1992 (first entry)
DE Pig brain thalamus polypeptide hormone fragment
KW Recombinant.

Wed Apr 28 09:08:27 1999

US-09-011-797-1.rag

Page 7

Db e e e e e e
Dt 788 argargargargargarg 810
QY 338 TGGAGCAGACGCTGCAGAA 360
Qt E Q K Q L Q K

Search completed: Fri Apr 16 14:34:06 1999
Job time : 304 secs.

THIS PAGE BLANK (USPTO)

Query Match 100.0%; Score 932; DB 34; Length 932;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 cctccagcgtgttcaagcagctgtcccgagctcctcaactctgcagagagagcttcac 60
1 cctccagcgtgttcaagcagctgtcccgagctcctcaactctgcagagagagcttcac 60
61 ccgagctcgggagagcttcaactgaagctgtgcatcctccagtgtagaagaagcttc 120
61 ccgagctcgggagagcttcaactgaagctgtgcatcctccagtgtagaagaagcttc 120
61 ccgagctcgggagagcttcaactgaagctgtgcatcctccagtgtagaagaagcttc 120
121 ccccgccctctctgactcttgcacaaagcagtgagcagtgactctgagcagctcagc 180
121 ccccgccctctctgactcttgcacaaagcagtgagcagtgactctgagcagctcagc 180
121 ccccgccctctctgactcttgcacaaagcagtgagcagtgactctgagcagctcagc 180
181 cctgctgagcagcagctcagctcagctcctcttaccagtcgaagagctcgagatgag 240
181 cctgctgagcagcagctcagctcagctcctcttaccagtcgaagagctcgagatgag 240
241 caccgtgaagagaatgcgcgctgtcagagatgtgtgtcgaagcccgagagcagagctgag 300
241 caccgtgaagagaatgcgcgctgtcagagatgtgtgtcgaagcccgagagcagagctgag 300
301 gcagatgcagagcctgtcagagatgcagatgcagatgcagatgcagatgcagatgcag 360
301 gcagatgcagagcctgtcagagatgcagatgcagatgcagatgcagatgcagatgcag 360
361 aggtttgggggcttcaactcagcagcagcagcagcagcagcagcagcagcagcagc 420
361 aggtttgggggcttcaactcagcagcagcagcagcagcagcagcagcagcagcagc 420
421 ttcagtgagttatgaagcagctactgtctgtcagatgcagatgcagatgcagatgcag 480
421 ttcagtgagttatgaagcagctactgtctgtcagatgcagatgcagatgcagatgcag 480
481 actctgcacagaaatgtaatgtgtagcagaagaagcagcagcagcagcagcagcagc 540
481 actctgcacagaaatgtaatgtgtagcagaagaagcagcagcagcagcagcagcagc 540
541 tgcagccatgaagatcagcagctgagcagcagcagcagcagcagcagcagcagcagc 600
541 tgcagccatgaagatcagcagctgagcagcagcagcagcagcagcagcagcagcagc 600
601 ggcgggagtcagagctcctcctcctgagcagcagcagcagcagcagcagcagcagc 660
601 ggcgggagtcagagctcctcctcctgagcagcagcagcagcagcagcagcagcagc 660
661 gcatgtcaccacaatcctgtgtgtagcagatgtagatgtagatgtagatgtagatgtag 720
661 gcatgtcaccacaatcctgtgtgtagcagatgtagatgtagatgtagatgtagatgtag 720
721 acatttaagggcccatcttctgcatcctcctcctcctcctcctcctcctcctcctcct 780
721 acatttaagggcccatcttctgcatcctcctcctcctcctcctcctcctcctcctcct 780
781 gaacatgaatcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
781 gaacatgaatcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
841 ctgtctcttctgactgtgtgtggaacaacttctccattaaacttactactagcaaaa 900
841 ctgtctcttctgactgtgtgtggaacaacttctccattaaacttactactagcaaaa 900
901 tggtaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 960
901 tggtaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 960
901 tggtaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 960
901 tggtaataaataaataaataaataaataaataaataaataaataaataaataaataaataa 960

```

RESULT 2
 ID Q76890 standard; DNA; 186 BP.
 AC Q76890;
 DT 23-SEP-1994 (first entry)

DE Human genome fragment.
 KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
 KW detection; homology; human; adrenal tissue; ds.
 OS Homo sapiens.
 PN M09401346-A.
 PF 20-JAN-1994.
 PF 13-JUL-1993; G01467.
 PR 13-JUL-1992; GB-014857.
 PA (MEDIC.) MEDICAL RES COUNCIL.
 PI Gross J, Hatfield KM, Howells D, Kelly M, Shaw D;
 DR WPI: 94-035056/04.
 PT New nucleic acid fragment encoding gene products - can be used
 PT for genetic analysis and mapping
 PS Claim 1; Page 256; 616pp; English.
 CC Human nucleic acid fragments, isolated from brain adrenal tissue,
 CC the placenta or bone marrow comprise any of: (A) a sequence
 CC selected from (Q76401-Q77613), (B) an allelic variation of a
 CC sequence as described in (A), or (C) a sequence complementary
 CC to (A) or (B).
 SQ Sequence 186 BP; 36 A; 63 C; 56 G; 31 T;

Query Match 12.2%; Score 114; DB 10; Length 186;
 Best Local Similarity 81.3%; Pred. No. 6.28e-45;
 Matches 148; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

4 aacctgaggtgcgcacctcctcagatgcagagagagcttcccgagccctctgagct 63
79 MACCTAAGCTGTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 138
64 ccatgcacaaagatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 123
139 CTTTGCACCAAGCCATGCGCAGTGTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAG 198
124 gtggcgctgctctcctcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
199 ACCTCCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 258
184 cg 185
259 CG 260

```

RESULT 3
 ID Q51746 standard; cDNA; 91 BP.
 AC Q51746;
 DT 31-MAY-1994 (first entry)
 DE Oligonucleotide probe MK14-A
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
 KW ss.
 OS Synthetic.
 PN EP-571911-A.
 PD 01-DEC-1993.
 PF 24-MAY-1993; 108325.
 PR 26-MAY-1992; US-889651.
 PA (BECT.) BECTON DICKINSON CO.
 PI Shank DD, Spears PA;
 DR WPI: 93-378844/48.
 PT New oligo:nucleotide probes specific for Mycobacteria - used for
 PT detection and amplification of Mycobacteria nucleic acid in
 PT samples
 PS Claim 3; Page 14; 23pp; English.
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
 CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
 CC cross reacted to a few non-mycobacterial spp. The probe may
 CC be useful as an initial screen for mycobacterial infection.
 CC See also Q51735-45 and Q51747-59.
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 4.6%; Score 43; DB 9; Length 91;
 Best Local Similarity 9.1%; Pred. No. 3.67e-08;
 Matches 5; Conservative 44; Mismatches 6; Indels 0; Gaps 0;

18-AUG-1994.
01-FEB-1994; U00977.
01-FEB-1993; US-013416.
30-DEC-1993; US-176500.
31-JAN-1994; US-189331.
(UYNC-) UNIV NORTH CAROLINA.
Fowlkes DM, Kay BK;
MPL: 94-279739/34.
P-PSDB: R65154.
Identifying proteins or peptide(s) which bind a ligand - by
screening a recombinant vector library expressing fusion proteins
comprising a binding domain and an effector domain
Disclouse: Page 35; 25pp; English.
070468 is a generic DNA sequence used to generate random TSAR (Totally
Synthetic Affinity Reagents) peptides. This generic formula can also be
represented as follows: X(NNB)11(TGC)(NNB)62(NNB)7(TGC)(NNB)10Y. X
and Y are flanking restriction sites (X is not the same as Y) that are
not specified further. Other generic sequences are shown in 070466-68.
Other specific peptides generated by these generic sequences are shown in
R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
comprising at least two functional regions - a binding domain with
affinity for a ligand and a second effector peptide portion that is
chemically or biologically active. They may further comprise a linker
peptide between the 2 domains. The oligonucleotides are also designed so
that the expressed peptide contains 2 or 4 cysteine residues positioned
in, or flanking, the unpredicted or variant residues. These residues
confer some degree of conformational rigidity to the peptides. The TSARs
or compns. comprising a TSAR binding domain can be used *in vivo* to
deliver a chemically or biologically active moiety, eg. metal ion.

18-AUG-1994.
01-FEB-1994; U00977.
01-FEB-1993; US-013416.
30-DEC-1993; US-176500.
31-JAN-1994; US-189331.
(UNNC-) UNIV NORTH CAROLINA.
Fowlkes DM, Kay BK;
WPI: 94-279739/34.
P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PT Disclosure; Page 35; 253pp; English.
Q70467 is a generic DNA sequence used to generate random TSAR (Totally
Synthetic Affinity Reagents) peptides. This generic formula can also be
represented as follows: X(NNN)16(TGC)(NNB)16(TGC)(NNB)11.Y. X
and Y are flanking restriction sites (X is not the same as Y) that are
not specified further. Other generic sequences are shown in Q70466-68.
R65151-55. TSARs are concatenated heterofunctional proteins or peptides,
comprising at least two functional regions - a binding domain with
affinity for a ligand and a second effector region. A binding domain that is
chemically or biologically active. They may further comprise a linker
peptide between the 2 domains. The oligonucleotides are also designed so
that the expressed peptide contains 2 or 4 cysteine residues positioned
in, or flanking, the unpredicted or variant residues. These residues
confer some degree of conformational rigidity to the peptides. The TSARs
or compms. comprising a TSAR binding domain can be used in vivo to
deliver a chemically or biologically active moiety, eg. metal ion,
radioisotope, peptide, toxin or enzyme, to the specific target or on the
cell. They can also replace the function of macromolecules, eg.
monoclonal or polyclonal antibodies and therefore circumvent the need for
complex methods of hybridoma formation or in vivo antibody production.
The TSARs are easily characterised and have designed activity allowing
direct and rapid detection. Sequence 118BP; 0 A; 2 C; 2 G; 2 T;
Sequence 118BP; 0 A; 2 C; 2 G; 2 T;

Cp 340 CAACTCATGCGCTCATCTGCGACAGAGCTCTGCATCTGCGTACAGGCTCTGCGTCTCGG 261
 Db 63 bnnbtycnbn 111
 Cp 280 CTTCGACACACACTCTGCACGCGGCACTTCTCTTCAGTGTCTGCAGTCTC 232
 RESULT 15
 ID Q70469 standard; DNA; 114 BP.
 AC Q70469;
 DT 07-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 FS key
 FH Location/Qualifiers
 FT misc_feature 55..60
 FT /tag= a
 FT /note= "this sequence represents '2'; Z can be a
 FT sequence of 6,9 or 12 nucleotides (see
 FT comments)"
 PN MO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UNNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PS comprising a binding domain and an effector domain
 PS Disclosure: Page 35: 255pp; English.
 PD Q70469 is a generic DNA sequence used to generate random TSAR peptide
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-
 CC (TGC)(NNB)6(XNNB)(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. This
 CC sequence generates peptides that are cleavable in structure. Other
 CC generic sequences are shown in Q70465-68. Other specific peptides
 CC generated by these generic sequences are shown in K65150-54. TSARs are
 CC concatenated heterofunctional proteins or peptides, comprising at least
 CC two functional regions - a binding domain with affinity for a ligand and
 CC a second effector peptide portion that is chemically or biologically
 CC active. They may further comprise a linker peptide between the 2 domains.
 CC The oligonucleotides are also designed so that the expressed peptide
 CC contains 2 or 4 cysteine residues positioned in, or flanking, the
 CC unpredicted or variant residues. These residues confer some degree of
 CC conformational rigidity to the peptides. The TSARs or compns. comprisn
 CC a TSAR binding domain can be used in vivo to deliver a chemically or
 CC biologically active moiety, eg. metal ion, radiolotope, peptide, toxin
 CC or enzyme, to the specific target or on the cell. They can also replace
 CC the function of macromolecules, eg. monoclonal or polyclonal antibodies
 CC and therefore circumvent the need for complex methods of hybridoma
 CC formation or in vivo antibody production. The TSARs are easily
 CC characterised and have designed activity allowing direct and rapid
 CC detection in a screening process.
 CC Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
 SQ
 Query Match 3.6%; Score 34; DB 12; Length 114;
 Best Local Similarity 4.7%; Pred. No. 4.97e-04;
 Matches 5; Conservative 31; Mismatches 70; Indels 0; Gaps 0;
 Db 6 bnnbnnbn 65
 QY 15 CAGCAGCTGTCGCCGAGAGACTGCTCACCCTCCAGGAGAGGCTCCACCGGCTCGG6CAG 74
 Db 66 btgcnbn 111
 QY 75 CTTCAACCTGAAGCTGTGCATCTCTCCAGTGTGAAGAGAAAGCTTTC 120

Wed Apr 28 09:08:29 1999

US-09-011-797-1.rmg

Page 8

Search completed: Tue Apr 27 11:29:51 1999
Job time : 154 secs.

 M I T O C H O N D R I A L
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:18:44 1999; MasPar time 2.64 Seconds
 Tabular output not generated. 173.116 Million cell updates/sec

Title: >US-09-011-797-2
 Description: (1-17) from US09011797.pep
 Perfect Score: 116
 Sequence: 1 FGFTGARKSARKLANQ 17

Scoring table:
 PAM 150
 GAP 15

Searched: 74019 segs, 26840295 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 27.528; Variance 32.859; scale 0.838

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	116	100.0	17	1	PNOC_PIG	3.09e-15
2	116	100.0	176	1	NOCICEPTIN (ORPHANIN F	3.09e-15
3	116	100.0	181	1	PNOC_HUMAN	3.09e-15
4	116	100.0	181	1	PNOC_RAT	3.09e-15
5	67	57.8	205	1	PNOC_MOUSE	3.09e-15
6	66	56.9	205	1	PNOC_MOUSE	3.09e-15
7	61	52.6	205	1	PNOC_MOUSE	3.09e-15
8	60	51.7	205	1	PNOC_MOUSE	3.09e-15
9	60	51.7	205	1	PNOC_MOUSE	3.09e-15
10	60	51.7	205	1	PNOC_MOUSE	3.09e-15
11	59	50.9	205	1	PNOC_MOUSE	3.09e-15
12	58	50.0	205	1	PNOC_MOUSE	3.09e-15
13	58	50.0	205	1	PNOC_MOUSE	3.09e-15
14	58	50.0	205	1	PNOC_MOUSE	3.09e-15
15	57	49.1	205	1	PNOC_MOUSE	3.09e-15
16	57	49.1	205	1	PNOC_MOUSE	3.09e-15
17	57	49.1	205	1	PNOC_MOUSE	3.09e-15
18	56	48.3	205	1	PNOC_MOUSE	3.09e-15
19	55	47.4	205	1	PNOC_MOUSE	3.09e-15
20	54	46.6	205	1	PNOC_MOUSE	3.09e-15
21	54	46.6	205	1	PNOC_MOUSE	3.09e-15
22	54	46.6	205	1	PNOC_MOUSE	3.09e-15
23	53	45.7	205	1	PNOC_MOUSE	3.09e-15

Result ID	Score	Query Match	Length DB	ID	Description	Pred. No.
24	53	45.7	93	1	RT19_MARPO	1.20e+01
25	53	45.7	141	1	Y221_MYCPN	1.20e+01
26	53	45.7	300	1	PEPM_TETPY	1.20e+01
27	53	45.7	323	1	DHAS_MYCBO	1.20e+01
28	53	45.7	388	1	METB_MYCLE	1.20e+01
29	53	45.7	476	1	EIBL_ADE40	1.20e+01
30	53	45.7	547	1	SAPA_ECOLI	1.20e+01
31	53	45.7	638	1	SYC_HUMAN	1.20e+01
32	53	45.7	692	1	YQIR_BACSU	1.20e+01
33	53	45.7	704	1	YBY1_YEAST	1.20e+01
34	52	44.8	90	1	DBH_RHIME	1.85e+01
35	52	44.8	125	1	MERF_STRLI	1.85e+01
36	52	44.8	205	1	RS4_HAELN	1.85e+01
37	52	44.8	305	1	NUBM_HUMAN	1.85e+01
38	52	44.8	330	1	KDGT_BACSU	1.85e+01
39	52	44.8	370	1	EXOH_RHIME	1.85e+01
40	52	44.8	394	1	XYLA_ARST7	1.85e+01
41	52	44.8	420	1	YS8_MYCPN	1.85e+01
42	52	44.8	430	1	YS8_MYCPN	1.85e+01
43	52	44.8	474	1	Y294_MYCGE	1.85e+01
44	52	44.8	1139	1	MA2X_HUMAN	1.85e+01
45	52	44.8	1597	1	SOL_DROME	1.85e+01

ALIGNMENTS

Result ID	Score	Query Match	Length DB	ID	Description	Pred. No.
AC	1	100.0%	17	AA.	PNOC_PIG	3.09e-15
AC	1	100.0%	17	AA.	PNOC_PIG	3.09e-15
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)				
DE	NOCICEPTIN (ORPHANIN FQ).					
GN	PNOC.					
OS	SUS SCROFA (PIG).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;					
OC	EUTHERIA; ARTIODACTYLA.					
RN	(1)					
RP	SEQUENCE.					
RC	TISSUE-HYPOTHALAMUS;					
RX	MEDLINE: 96055113					
RA	REINSCHEID R.K., NOTACKER H.-P., BOURSON A., ARDATI A.,					
RA	HENNINGSSEN R.A., BUNZOW J.R., GRANDY D.K., LANGEN H., MONSMA F.J. JR.,					
RA	GIVELLI O.;					
RL	SCIENCE 270:792-794(1995).					
CC	- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE					
CC	RECEPTOR (OPR1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY					
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED IN					
CC	NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED					
CC	INTRACEREBROVENTRICULARLY INTO MICE, THIS PEPTIDE INDUCES					
CC	HYPALGESIA AND DECREASES LOCOMOTOR ACTIVITY.					
CC	- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.					
CC	- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS					
CC	FAMILY.					
DR	PROSITE: PS01252; OPIOIDS PRECURSOR; PARTIAL.					
KW	OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE.					
SEQ	SEQUENCE 17 AA: 1809 MW; 09F05DDE CRC32;					
Query Match	100.0%; Score 116; DB 1; Length 17;					
Best local Similarity	100.0%; Pred. No. 3.09e-15;					
Matches	17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Db	1 FGFTGARKSARKLANQ 17					
QY	1 FGFTGARKSARKLANQ 17					
RESULT	2					
ID	PNOC_HUMAN	STANDARD:	PRT:	176 AA.		
AC	013519;					
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)					
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)					

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC).
 GN PNOC OR OFQ.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; PRIMATES.
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 MEUNIER J.-C., PARMENTIER M.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA NOTHACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSEN R.A.,
 MONSMA F.J. JR., WATSON S.J., CIVELLI O.;
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
 RN [3]
 RP STRUCTURE BY NMR OF NOCICEPTIN.
 RX MEDLINE: 97312464.
 RA SALVADORI S., PICONE D., TANCREDI T., GUERRINI R., SPADACCINI R.,
 LAZARUS L.H., REGOLI D., TEMUSSI P.A.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 233:640-643(1997).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 DR EMBL: X97370; E244777; -;
 DR EMBL: X97367; E244785; -;
 DR EMBL: X97368; E244785; JOINED.
 DR EMBL: U48263; G1185010; -;
 DR MIM: 601459; -;
 DR PROSITE: PS01252; OPIOIDS PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 19
 FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
 SQ SEQUENCE 176 AA; 20295 MW; 8055843A CRC32;
 Query Match 100.0%; Score 116; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 3.09e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 130 FGGFTGARRSARRLANQ 146
 QY 1 FGGFTGARRSARRLANQ 17
 RESULT 3
 ID PNOC-RAT STANDARD: PRT; 181 AA.
 AC 062923; 064162;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC) (ORL1 RECEPTOR AGONIST).
 GN PNOC.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96216718.
 RA HOTTANT T., NISHI M., TAKESHITA H., NUKADA T., SUGIMOTO T.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP STRAIN-SPRAGUE-DAWLEY;
 RN

RA NOTHACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSEN R.A.,
 MONSMA F.J. JR., WATSON S.J., CIVELLI O.;
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 MEUNIER J.-C., PARMENTIER M.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [3]
 RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
 RC TISSUE-BRAIN;
 RX MEDLINE: 96013745.
 RA MEUNIER J.-C., MOLLEREAU C., TOLL L., SUAUDEAU C., MOISAND C.,
 ALVIERIE P., BOUTON J.-L., GUILLEMOT J.-C., FERRARA P.,
 MONSARRAT B., MAZARQUI H., VASSART G., PARMENTIER M., COSENTIN J.;
 RL NATURE 377:532-535(1995).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
 AND BRAIN, BEING MORE ABUNDANT IN THE HYPOTHALAMUS AND STRIATUM.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 DR EMBL: U48262; G1185012; -;
 DR EMBL: X97375; E244795; -;
 DR EMBL: S79730; G1172239; -;
 DR PROSITE: PS01252; OPIOIDS PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
 FT DOMAIN 109 120 2 X 6 AA TANDEM REPEATS OF D-A-E-P-X-A.
 FT REPEAT 109 114 1.
 FT REPEAT 115 120 2.
 SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;
 Query Match 100.0%; Score 116; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 3.09e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 135 FGGFTGARRSARRLANQ 151
 QY 1 FGGFTGARRSARRLANQ 17
 RESULT 4
 ID PNOC-MOUSE STANDARD: PRT; 187 AA.
 AC 064387; 061105; 061938;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC) (N23K / N27K).
 GN PNOC OR NNC1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 96216718.
 RA HOTTANT T., NISHI M., TAKESHITA H., NUKADA T., SUGIMOTO T.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]

KM RIBOSOMAL PROTEIN; RNA-BINDING.
SQ SEQUENCE 205 AA; 23947 MW; 02D16821 CRC32;

Query Match 56.9%; Score 66; DB 1; Length 205;
Best Local Similarity 53.3%; Pred. No. 2 41e-02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPRKSRKRVNH 119
||:|||||:|
QY 3 GFTGARKSRKRLAQ 17

RESULT 7
ID G3PA_PEA STANDARD; PRT; 405 AA.
AC P12858;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A PRECURSOR, CHLOROPLAST
DE (EC 1.2.1.12).
GN GAPA OR GPAL.
OS PISUM SATIVUM (GARDEN PEA).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
OC FABACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. ROSAKRONE; TISSUE-SEEDLING;
RX LIAUD M.-F., ZHANG D.-X., CERRE R.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:8918-8922(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. ROSAKRONE; TISSUE-SEEDLING;
RX MEDLINE; 93357435.
RL BRINKMANN H., CERRE R., SALOMON M., SOLI J.;
PLANT MOL. BIOL. 13:81-94(1989).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
+ NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: CALVIN CYCLE.
CC -1- SUBUNIT: TRIMER OF EITHER FOUR A CHAINS (GAPDH 2) OR TWO A AND
TWO B CHAINS (GAPDH 1).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- PLANTS CONTAIN THREE FORMS OF GAPDH: A CYTOSOLIC FORM WHICH
PARTICIPATES IN PHOTOSYNTHESIS AND TWO CHLOROPLAST FORMS WHICH
PARTICIPATES IN PHOTOSYNTHESIS. THESE THREE FORMS ARE ENCODED BY
DISTINCT GENES.
DR EMBL; X52148; G12159; -;
DR EMBL; X15190; G20729; -;
DR PIR; S14243; DEPMNA.
DR HSSP; P00362; 1GDI.
DR PROSITE; PS00071; GAPDH. 1.
KW CALVIN CYCLE; OXIDOREDUCTASE; NAD; TRANSIT PEPTIDE; CHLOROPLAST;
KM MULTIGENE FAMILY.
FT TRANSIT 1 68
FT CHAIN 69 405
FT FT
FT BINDING 222 222
FT ACT_SITE 249 249
FT CONFLICT 194 194
SQ SEQUENCE 405 AA; 43338 MW; A9F53CE6 CRC32;
G -> R (IN REF. 2).

Query Match 52.6%; Score 61; DB 1; Length 405;
Best Local Similarity 46.7%; Pred. No. 2 94e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 20 FSEFGLRNSRHLP 34
|:|:|:|:|:|:|
QY 1 FCGTGAARKSRKRLA 15

RESULT 8
ID PNAD_PIG STANDARD; PRT; 310 AA.
AC Q28955;

DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN N-TERMINAL ASPARAGINE AMIDOHYDROLASE (EC 3.5.1.-) (PROTEIN
DE NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN-AMIDASE) (PNAD).
GN NTNL.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.
OC EUTHERIA; ARTIODACTYLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.
RX MEDLINE; 95113832.
RA STEWART A.E., ARFIN S.M., BRADSHAW R.A.;
RL J. BIOL. CHEM. 270:25-28(1995).
RN [2]
RP CHARACTERIZATION.
RC TISSUE-LIVER.
RX MEDLINE; 94375451.
RA STEWART A.E., ARFIN S.M., BRADSHAW R.A.;
RL J. BIOL. CHEM. 269:23509-23517(1994).
CC -1- FUNCTION: SIDE CHAIN DEAMIDATION OF N-TERMINAL ASPARAGINE RESIDUES
TO ASPARTIC ACID. REQUIRED FOR THE UBQUITIN-DEPENDENT TURNOVER OF
INTRACELLULAR PROTEINS THAT INITIATE WITH MET-ASN. THESE PROTEINS
ARE ACETYLATED ON THE RETAINED INITIATOR METHIONINE AND CAN
SUBSEQUENTLY BE MODIFIED BY THE REMOVAL OF N-ACETYL METHIONINE BY
ACYLAMINOACID HYDROLASE. CONVERSION OF THE RESULTING NH2-TERMINAL
ASPARAGINE TO ASPARTIC ACID BY PNAD RENDERS THE PROTEIN
SUSCEPTIBLE TO ARGINYLAATION, POLYUBQUITINYLAATION AND DEGRADATION
AS SPECIFIED BY THE N-END RULE.
CC -1- SUBUNIT: MONOMER.
DR EMBL; U17062; G595950; -;
KW HYDROLASE.
SQ SEQUENCE 310 AA; 34760 MW; 3110E5ED CRC32;

Query Match 51.7%; Score 60; DB 1; Length 310;
Best Local Similarity 43.8%; Pred. No. 4 77e-01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 127 GGFSDRQLSKRLTHQ 142
||:|:|:|:|:|
QY 2 GGFSGARKSRKRLAQ 17

RESULT 9
ID PNAD_MOUSE STANDARD; PRT; 310 AA.
AC Q64311;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN N-TERMINAL ASPARAGINE AMIDOHYDROLASE (EC 3.5.1.-) (PROTEIN
DE NH2-TERMINAL ASPARAGINE DEAMIDASE) (NTN-AMIDASE) (PNAD).
GN NTNL.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C129.
RX MEDLINE; 97067079.
RA GRIGOREV S., STEWART A.E., KWON Y.T., ARFIN S.M., BRADSHAW R.A.,
RA JERKINS N.A., COPELAND N.G., VARSHAVSKY A.;
RL J. BIOL. CHEM. 271:28521-28532(1996).
CC -1- FUNCTION: SIDE CHAIN DEAMIDATION OF N-TERMINAL ASPARAGINE RESIDUES
TO ASPARTIC ACID. REQUIRED FOR THE UBQUITIN-DEPENDENT TURNOVER OF
INTRACELLULAR PROTEINS THAT INITIATE WITH MET-ASN. THESE PROTEINS
ARE ACETYLATED ON THE RETAINED INITIATOR METHIONINE AND CAN
SUBSEQUENTLY BE MODIFIED BY THE REMOVAL OF N-ACETYL METHIONINE BY
ACYLAMINOACID HYDROLASE. CONVERSION OF THE RESULTING NH2-TERMINAL
ASPARAGINE TO ASPARTIC ACID BY PNAD RENDERS THE PROTEIN
SUSCEPTIBLE TO ARGINYLAATION, POLYUBQUITINYLAATION AND DEGRADATION
AS SPECIFIED BY THE N-END RULE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).

DR EMBL: U57692; G1373367; -
 DR EMBL: U57691; G1373365; -
 DR MGI: 108471; NTANTL.
 KW HYDROLASE.
 SQ SEQUENCE 310 AA; 34595 MW; 739412CE CRC32;

Query Match 51.7%; Score 60; DB 1; Length 310;
 Best Local Similarity 43.8%; Pred. No. 4.77e-01;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 127 GGFSDRQLSOKLTHQ 142
 |||:|::|::|:
 QY 2 GGFCAKRSARKLANQ 17

RESULT 10
 ID CUL1_HUMAN STANDARD: PRT: 752 AA.
 AC 013616;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE CULLIN HOMOLOG 1 (CUL-1).
 GN CUL1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96279828.
 RA KIREPOS E.T., LANDER L.E., WING J.P., HE W.W., HEDGECOCK E.M.;
 RL CELL 85:839-839(1996).
 CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
 DR EMBL: U58087; G1381142; -
 DR PROSITE: PS01256; CULLIN_1; 1.
 DR PROSITE: PS0069; CULLIN_2; 1.
 SQ SEQUENCE 752 AA; 87387 MW; 6CEB00D8 CRC32;

Query Match 51.7%; Score 60; DB 1; Length 752;
 Best Local Similarity 46.7%; Pred. No. 4.77e-01;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 542 FTAFYASRSGRKL 556
 |::|::|::|:
 QY 1 FGFTGAKRSARKLA 15

RESULT 11
 ID L119_DROME STANDARD: PRT: 773 AA.
 AC 024311;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE LIN-19 HOMOLOG PROTEIN.
 GN LIN19.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R; TISSUE-BRAIN;
 RA FILIPPOV V.A., FILIPPOVA M.A., SEHNAL F.;
 RL SUBMITTERED (JAN-1996) TO EMBL/GENBANK/DBS DATA BANKS.
 CC -1- SIMILARITY: BELONGS TO THE CULLIN FAMILY.
 DR EMBL: L41642; G1146334; -
 DR FLYBASE: FB990015509; lin19.
 DR PROSITE: PS01256; CULLIN_1; 1.
 DR PROSITE: PS0069; CULLIN_2; 1.
 SQ SEQUENCE 773 AA; 89411 MW; 8851D23F CRC32;

Query Match 50.9%; Score 59; DB 1; Length 773;
 Best Local Similarity 57.1%; Pred. No. 7.70e-01;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 565 FNEFYAHSGRKL 578

QY 1 FGFTGAKRSARKLA 14
 |::|::|::|:

RESULT 12
 ID RS4_HELPY STANDARD: PRT: 208 AA.
 AC P56011;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE 30S RIBOSOMAL PROTEIN S4.
 GN RPSD OR HP1294
 OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
 OC AEROBIC; MOTILE; HELICAL AND/OR VIBRIOID.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALIL H.G., GLODER A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCARNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RL NATURE 388:539-547(1997).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL: AE000633; G2314460; -
 DR TIGR: HP1294; -
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 KW RIBOSOMAL PROTEIN; RRNA-BINDING.
 SQ SEQUENCE 208 AA; 23964 MW; EE21B30C CRC32;

Query Match 50.0%; Score 58; DB 1; Length 208;
 Best Local Similarity 46.7%; Pred. No. 1.24e+00;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 108 GFATRSSARQLVTH 122
 ||::|::|::|:
 QY 3 GFTGAKRSARKLANQ 17

RESULT 13
 ID NASC_BACSU STANDARD: PRT: 421 AA.
 AC P42432;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NITRATE TRANSPORTER.
 GN NASC.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 95173124.
 RA OGAWA K.-I., AKAGAWA E., YAMANE K., SUN Z.-W., LACELLE M., ZUBER P.,
 RA NAKANO M.M.;
 RL J. BACTERIOL. 177:1409-1413(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 97124189.
 RA YAMANE K., KUNANO M., KURITA K.;
 RL MICROBIOLOGY 142:3047-3056(1996).
 CC -1- FUNCTION: MAY FUNCTION AS A NITRATE TRANSPORTER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE NARX/NASA FAMILY OF TRANSPORTERS.

DR EMBL: D30689; G1437473; -
 DR EMBL: D50453; G1805405; -
 DR EMBL: 299105; E1182285; -
 DR SUBLIST: EG11093; NASA.
 KW TRANSPORT; TRANSMEMBRANE; NITRATE ASSIMILATION.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 79 99 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 FT TRANSMEM 167 187 POTENTIAL.
 FT TRANSMEM 217 237 POTENTIAL.
 FT TRANSMEM 248 268 POTENTIAL.
 FT TRANSMEM 277 297 POTENTIAL.
 SQ SEQUENCE 421 AA; 46067 MW; D7B4779 CRC32;
 Query Match
 Best Local Similarity 50.0%; Score 58; DB 1; Length 421;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 DB 388 GKGARKARRM 400
 QY 2 GGFTGARKARKL 14

RESULT 14
 ID CBF_HUMAN STANDARD; PRT; 998 AA.
 AC 003701;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DE CCAAT-BINDING FACTOR (CBF).
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91061780.
 RA LUM L., SOLTZMAN L., KAUFMAN R., LINZER D.I.H., WU B.;
 RL MOL. CELL. BIOL. 10:6709-6717(1990).
 CC -1- FUNCTION: STIMULATES TRANSCRIPTION FROM THE HSP70 PROMOTER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 DR EMBL: M37197; G179969; -
 KW TRANSCRIPTION REGULATION; ACTIVATOR; NUCLEAR PROTEIN.
 SQ SEQUENCE 998 AA; 114071 MW; 6DD27B10 CRC32;

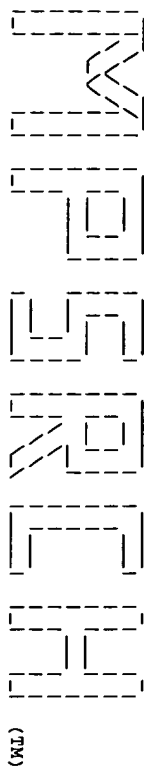
Query Match
 Best Local Similarity 50.0%; Score 58; DB 1; Length 998;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 958 GSFOGPRKKRNND 972
 QY 2 GGFTGARKARKL 16

RESULT 15
 ID YICN_ECOLI STANDARD; PRT; 159 AA.
 AC P31439;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE HYPOTHETICAL 18.2 KD PROTEIN IN NUPA-UHPT INTERGENIC REGION.
 GN YICN
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE; 93315143.
 RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
 RL GENOMICS 16:551-561(1993).
 DR EMBL: L10328; G290512; -

DR EMBL: AE000444; G1790096; -
 DR ECGENE: EG11690; YICN.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 159 AA; 18173 MW; F09A1692 CRC32;
 Query Match
 Best Local Similarity 49.1%; Score 57; DB 1; Length 159;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 DB 31 SGARKAIRLSDR 43
 QY 5 TGARKSARKLANQ 17

Search completed: Fri Apr 16 14:18:52 1999
 Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_tpm n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:21:16 1999; Maspar time 137.97 Seconds

Tabular output not generated. 873.545 Million cell updates/sec

Title: >US-09-011-797-3
Description: (1-17) from US09011797.pep
Perfect Score: 215
N.A. Sequence: 1 TTYWSNGARTTATGNGMCA.....TNSNATGACARMSWNSCAR 51
Comp: AARMSNCTVAAFPACKNGT.....ANMSNACAGYMSWNSNGTY

Scoring table: TABLE bkttranslated2
Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_v1

Database:

genbank110
16:gb_ba1 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl3
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v1

Statistics: Mean 48.293; Variance 156.627; scale 0.308

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	199	92.6	372	29	MMU44027	Mus musculus orphanin
2	199	92.6	633	29	KMPPMNRNA	M.musculus mRNA for pr
3	199	92.6	878	26	HSPPNEX3	H.sapiens gene encodin
4	199	92.6	925	29	S79730	ORL1 receptor agonist
5	199	92.6	972	29	MUSN27K	Mouse N23K mRNA for de
6	199	92.6	972	29	RNU48262	Rattus norvegicus pre-
7	199	92.6	988	29	KMPPMNRNA	R.norvegicus mRNA for
8	199	92.6	1015	26	HSPPNMRNA	H.sapiens mRNA for pre
9	199	92.6	1047	29	MUSN23K	Mouse N23K mRNA for N2
10	199	92.6	1198	28	HSU48263	Human pre-pro-orphanin
11	199	92.6	1209	20	AB005251	Bos taurus mRNA for no
12	199	92.6	1253	29	KMPPNEX3	M.musculus gene encodi

13	199	92.6	1354	29	MUSNOP	Mouse mRNA for nocicep
14	124	57.7	2940	95	HS1170K4	Human DNA sequence ***
15	119	55.3	1669	83	AC002541	Human BAC clone RG043K
16	118	54.9	329	81	CE1ZK418	Caenorhabditis elegans
17	118	54.9	745	89	AB010068	Arabidopsis thaliana g
18	117	54.4	1562	37	HS111B22	Human DNA sequence ***
19	116	54.0	698	22	AC003121	Drosophila melanogaste
20	114	53.0	415	52	CEC17D12	Caenorhabditis elegans
21	114	53.0	1183	57	HS1142C11	Human DNA sequence fro
22	113	52.6	1123	24	NT33K0EP2	N.tabacum mRNA for 33
23	113	52.6	1175	24	LE33K0ECP	L.esculentum mRNA for
24	113	52.6	2901	6	CET23B5	Caenorhabditis elegans
25	113	52.6	952	67	HS91J24	Human DNA sequence ***
26	112	52.1	762	19	A1U25289	Aramigus intermedicus c
27	112	52.1	1006	35	AC003104	Homo sapiens chromosome
28	112	52.1	1414	60	AC004829	*** SEQUENCING IN PROG
29	111	51.6	1798	61	CE1ZK75	Caenorhabditis elegans
30	109	50.7	2757	29	MMGR5505	Mus musculus G protein
31	109	50.7	3048	19	CE22B3	Caenorhabditis elegans
32	109	50.7	7802	28	HS780M13	Human DNA sequence fro
33	109	50.7	8695	27	AC004464	Homo sapiens 12q13.1 p
34	109	50.7	1907	78	HS142L7	Homo sapiens DNA sequ
35	108	50.2	180	34	SEV6KPRO	Semliki Forest Virus (
36	108	50.2	3000	29	RNU14908	Rattus norvegicus cyto
37	108	50.2	384	34	SEV6K	Semliki Forest Virus R
38	108	50.2	1151	22	AR001293	Sequence 1 from patent
39	108	50.2	2615	4	CE33E2	Caenorhabditis elegans
40	108	50.2	3976	28	AC005853	Homo sapiens clone UMG
41	108	50.2	3980	25	AC002096	Homo sapiens chromosome
42	108	50.2	7773	27	HUAC004131	Homo sapiens chromosome
43	108	50.2	1259	87	HS306D1	Human DNA sequence fro
44	108	50.2	1444	74	AF051934	Homo sapiens chromosome
45	108	50.2	3480	50	HS229042	Homo sapiens 959 kb co

ALIGNMENTS

RESULT 1	MMU44027	372 bp	mRNA	ROD	30-MAY-1996
LOCUS	Mus musculus orphanin	FO/nociceptin precursor	mRNA, partial cds.		
DEFINITION	U44027				
ACCESSION	g1335869				
NID					
KEYWORDS					
SOURCE	house mouse.				
ORGANISM	Mus musculus.				
REFERENCE	1 (bases 1 to 372)				
AUTHORS	Pan,Y.X., Xu,J. and Pasternak,G.W.				
TITLE	Cloning and expression of a cDNA encoding a mouse brain orphanin				
JOURNAL	FO/nociceptin precursor				
MEDLINE	Biochem. J. 315 (Pt 1), 11-13 (1996)				
REFERENCE	2 (bases 1 to 372)				
AUTHORS	Pan,Y.-X.				
TITLE	Direct Submission				
JOURNAL	Submitted (28-DEC-1995) Ying-Xian Pan, Neurology, Memorial				
	Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY				
	10021, USA				
FEATURES	Location/Qualifiers				
source	1..372				
	/organism="Mus musculus"				
	/db_xref="taxon:10090"				
	/tissue_type="brain"				
	<2..331				
CDS	/codon_start=1				
	/product="orphanin FO/nociceptin precursor"				
	/db_xref="PID:g1335870"				
	/translation="TTSALYOPKASEMHTKMPVRSIVGVDAEPGADAPGADA				
	EPGADAEVEVKQLOKRGGFTGAKRSARLANKRSEFPRQIVLVLSMOSSQRRRT				
	LHONGNV"				
misc_feature	188..238				

Query Match	92.68;	Score 199;	DB 29;	Length 372,
Best Local Similarity	49.08;	Pred. No. 3.66e-07;		
Matches	25; Conservative	18; Mismatches	0	

9Y 1 TTYWSNGARTTYATGMCNCARTAYTYTNGTNTYMSNATGCARWSNWSNCAR 51

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae Mus.

REFERENCE AUTHORS	TITLE	JOURNAL
1 (bases 1 to 633) MollerEAU C., SIMONS, M. J., SOULARNE, P., LINERS, F., VASSART, G., MEUNIER, J. C., and PARMENTIER, M.	Structure, tissue distribution, and chromosomal localization of the preproreceptin gene	Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
96333281		

Bruxelles, I R I B H N ULB
Campus Erasme, 808 Route de Lennik, B
1070 Bruxelles, BELGIUM

1.633

```

mRNA      join(<1.      .150,151.      .>633)
CDS       join(25.    .150,151.    .588)

```

```

mRNA      join(<1.      .150,151.      .>633)
CDS       join(25.  .150,151.      .588)

```

```
/product="prepronociceptin"  
151. .>633
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match	92.6%;	Score 199;	DB 29;	Length 633;
Best Local Similarity	49.0%;	Pred. No. 3.66e-07;		
Matches	25; Conservative	18; Mismatches	8; Indels	

Db 502 TTCAGTGAATTATGAGGCAGTACCTGTCCTGAGCATGACGTCAACGCAA 55
|::: |:||::|: |::|: |::|: |::|: |::|: |::|: |::|: |::|: |::|:
QY 1 TTYSNCARTTYATGMCNCAATAYTTNCTNTMNSNAIGCARWMSWSCAR 51

RESULT	3	HSPNEX3	878 bp	DNA	PRI	10-SEP-1996
LOCUS		H.sapiens gene encoding	prepronociceptin, exon 3.			
DEFINITION		K97268				
ACCESSION		91331986				
NID		prepronociceptin.				
KEYWORDS		human.				
SOURCE						

REFERENCE
1 (bases 1 to 878)

TITLE Menier, J.C. and Parmetier, M.
JOURNAL Structure, tissue distribution, and chromosomal localization of the
MEDLINE preproreceptor gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
96323281

TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Université libre de

FEATURES	Location/Qualifiers
source	1. .878

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone_lib="lambda DASHII (Stratagene #94520)"  
/clone="HG91"
```

exor.

BASE COUNT	194 a	274 c	231 g	179 t
ORIGIN				

Query Match	92.6%;	Score 199;	DB 26;	Length 878;
Best Local Similarity	49.0%;	Pred. No. 3.66e-07;		
Matches	25;	Conservative	18.	Mismatches 8.

Dd 391 TTCAGCAGATTATGAGCAATACTGTGTCCTGAGCATGCACGACG 441
||::|:||:||||:|:|:|:| | : : |||||: : : ||:
Qy 1 TTYMSGARPTTATGMCNCATAYTTNGTVYTNWSNATGCARMSNMNCR 51

LOCUS	4	925 bp	ROD	30-JAN-1996
DEFINITION	S79730	mRNA		
ACCESSION	ORL1 receptor agonist precursor-endogenous agonist of opioid receptor-like ORL1 receptor [rats, brain, mRNA Partial, 925 nt]			
NID	S79730			
REVISION	g1172238			

SOURCE ORGANISM

REFERENCE 1 (bases 1 to 925)

Aounous
 Meunier, J. C.,
 Alvinerie, P.,
 Butour, J. L.,
 Mollereau, C.,
 Toll, L.,
 Snaudeau, C.,
 Moisan, C.,
 Guillemot, J. C.,
 Ferrara, P.,
 Monsarrat, B.

JOURNAL	Nature 377 (6549), 532-535 (1995)
MEDLINE	96013745
REMARK	GenBank staff at the National Center for

GenBank stair at the National Library of Medicine created this entry [NCBI g1bbsq 171655] from the original journal article.

OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYNSNATGCARMSWNSNCAR 51

RESULT 7
LOCUS RNPENRNA 988 bp RNA ROD 10-SEP-1996
DEFINITION R.norvegicus mRNA for prepronociceptin.
ACCESSION X97375
NID 91532064
KEYWORDS prepronociceptin.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 988)
AUTHORS Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (15), 8666-8670 (1996)
MEDLINE 96323281
REFERENCE 2 (bases 1 to 988)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite libre de
Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
COMMENT Related sequence 579730.
FEATURES
source location/Qualifiers
1..988
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_1lb="lambda DASHII (Stratagene #943501)"
/clone_1lb="lambda ZAPII (brain cDNA)"
/clone="RGE109"
/clone="RB101"
<1..150
/number=2
join(<1..150,151..615,616..974)
25..69
join(25..150,151..570)
/product="prepronociceptin"
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:6244795"
/db_xref="PID:91532065"
/translation="MKILFCVLLSLSSVSSGSCDCLTCOERLHPAPSPYLKIC
ILQCEKVPPLMTCTKAMASDESESPADPETLSALYOSKASEQHLKRPVR
SVYQADAEPEADAPVADEADDEVQKQKRFGGFTGARKSARKLANOKRSEFMQ
YVLMSOSQRRRLHNGNV"
70..567
/product="prepronociceptin"
151..615
/number=3
616..974
/number=4
971

BASE COUNT 231 a 292 c 250 g 215 t
ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 988;
Best Local Similarity 49.0%; Pred. No. 3,666-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 484 TTCAGTGGATTATAGGACGACTGCTGCTGACATGCAGTCAAGCCAA 534
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYNSNATGCARMSWNSNCAR 51

RESULT 8
LOCUS HSPENRNA 1015 bp RNA PRI 10-SEP-1996
DEFINITION H.sapiens mRNA for prepronociceptin.
ACCESSION X97370
NID 91531988

KEYWORDS prepronociceptin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1015)
AUTHORS Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (15), 8666-8670 (1996)
MEDLINE 96323281
REFERENCE 2 (bases 1 to 1015)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite libre de
Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
FEATURES
source location/Qualifiers
1..1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="lambda DASHII (Stratagene #94520)"
/clone="HG91"
/clone="RGE107"
/chromosome="8"
/map="p21"
/map="1.9 CR from WI-1172 marker"
<1..149
/number=2
join(<1..149,150..601,602..>1015)
24..68
join(24..149,150..554)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:6244777"
/db_xref="PID:91531989"
/translation="MKVLCDDLILSLSSVSSGSCDCLTCOERLHPALSDLEVC
ILECEKVPPLMTCTKAMASDESESPADPETLSALYOSKASEQHLKRPVR
SVYQADAEPEADAPVADEADDEVQKQKRFGGFTGARKSARKLANOKRSEFMQYLVIS
MOSORRRLHNGNV"
69..551
/product="prepronociceptin"
150..601
/number=3
602..>1015
/number=4
1015

BASE COUNT 212 a 304 c 264 g 235 t
ORIGIN

Query Match 92.6%; Score 199; DB 26; Length 1015;
Best Local Similarity 49.0%; Pred. No. 3,666-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 468 TTCAGTGGATTATAGGACGACTGCTGCTGACATGCAGTCAAGCCAG 518
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
OY 1 TTYWSNGARTTATGMCNCARTAYTNGTNTYNSNATGCARMSWNSNCAR 51

RESULT 9
LOCUS MUSEN23K 1047 bp RNA ROD 14-MAR-1997
DEFINITION Mouse N23K mRNA for N23K protein, complete cds.
ACCESSION D50056
NID 91857023
KEYWORDS N23K protein; N23K.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 1047)
AUTHORS Salto,Y., Maruyama,K., Saido,T.C. and Kawashima,S.

TITLE N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin

JOURNAL MEDLINE Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)

REFERENCE 2 (bases 1 to 1047)

AUTHORS Saito, Y.

JOURNAL Direct Submission

Submitted (05-APR-1995) to the DDBJ/EMBL/GenBank databases. Yumiko Saito, Tokyo Metropolitan Institute for Medical Science, molecular biology; Honkomagome 3-18-22, Bunkyo-Ku, Tokyo 113, Japan (E-mail: saito@rinsoken.or.jp, Tel: 03-3823-2101 (ex. 5248), Fax: 03-5685-6609)

COMMENT On Mar 1, 1997 this sequence version replaced gi:1217905. Sequence updated (06-Mar-1996) by: Yumiko Saito.

FEATURES Location/Qualifiers

source 1..1047

gene /organism="Mus musculus"

CDS /db_xref="taxon:10090"

/cell_line="NS20Y"

/cell_type="neuroblastoma"

57..620

/gene="N23K"

57..620

/note="neuropeptide precursor"

/product="N23K protein"

/db_xref="PID:d1009395"

/db_xref="PID:g1217906"

/translation="MKILFCDDVLLSLFSSVSSCORDCITCOEKLHPADSPNLTAC ILCEEKVPPSPPLMTPTCTRYMAGSSGSLSPADPELYSALIQKASEMOKLRMPYR SLVQVDAEPGADAEPADEPADEAEVQKQKQKRGFTGARSARKLANQKRF SEFMROYLVLSMSSQRRLHONGNV"

Db 534 TTCAGTAGTTATGAGCAGTACTGCTCTGATGCGATGCAAGTCA 584

LOCUS 10440..1045

BASE COUNT 237 a 315 c 263 g 232 t

ORIGIN

Query Match 92.6%; Score 199; DB 29; Length 1047;

Best Local Similarity 49.0%; Pred. No. 3,66e-07;

Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 1 TTWMSGARITTYATGMCNCAITAYTNGTNTWMSNATGCAKMSWNSCAR 51

RESULT 10

LOCUS HSU48263 1198 bp mRNA PRI 23-AUG-1996

DEFINITION Human pre-pro-orphanin FQ (OFQ) mRNA, complete cds.

ACCESSION U48263

NID 91185009

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Carnivora; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1198)

AUTHORS Nothacker, H.P., Reinscheid, R.K., Mansour, A., Henningsen, R.A., Ardati, A., Monama, F.J. Jr., Watson, S.J. and Civelli, O.

JOURNAL Primary structure and tissue distribution of the orphanin FQ precursor

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)

REFERENCE 2 (bases 1 to 1198)

AUTHORS Nothacker, H.-P. and Henningsen, R.A.

JOURNAL Direct Submission

Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202, Hoffmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland

FEATURES Location/Qualifiers

source 1..1198

/organism="Homo sapiens"

/db_xref="taxon:9606"

5'UTR /tissue_type="brain"

gene 1..211

212..742

/gene="OFQ"

212..742

/gene="OFQ"

212..742

/codon_start=1

/product="pre-pro-orphanin FQ"

/db_xref="PID:g1185010"

/translation="MKVLLCDLLSLFSSVSSCORDCITCOEKLHPADSPLEVC ILCEEKVPPSPPLMTPTCTRYMAGSSGSLSPADPELYSALIQKASEMOKLRMPYR SLVQVDAEPGADAEPADEPADEAEVQKQKQKRGFTGARSARKLANQKRFSEFMROYLVLS MSSQRRLHONGNV"

sig_peptide 212..268

mat_peptide /gene="OFQ"

539..649

/gene="OFQ"

/product="orphanin FQ"

743..1198

3'UTR 1198

polya_site /note="29 A nucleotides"

BASE COUNT 244 a 335 c 345 g 273 t 1 others

ORIGIN

Query Match 92.6%; Score 199; DB 28; Length 1198;

Best Local Similarity 49.0%; Pred. No. 3,66e-07;

Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 656 TTCAGTAGTTATGAGCAGTACTGCTCTGATGCGATGCAAGTCA 706

LOCUS 11

DEFINITION Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds.

ACCESSION AB005251

NID 93041768

KEYWORDS nociceptin/orphanin FQ precursor.

SOURCE Bos taurus CDNA to mRNA.

ORGANISM Bos taurus

REFERENCE 1 (sites)

AUTHORS Okuda-Ashitaka, E., Minami, T., Tachibana, S., Yoshihara, Y., Nishitani, Y., Kimura, T. and Ito, S.

JOURNAL Nocistatin, a peptide that blocks nociceptin action in pain transmission

MEDLINE Nature 392 (6673), 286-289 (1998)

REFERENCE 2 (bases 1 to 1209)

AUTHORS Okuda-Ashitaka, E.

JOURNAL Direct Submission

Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Okuda-Ashitaka, Kansai Medical University, Department of Medical Chemistry, 10-15 Fumazono, Moriyuchi, Osaka, JAPAN, Moriyuchi, Osaka 570, Japan (E-mail: ashitaka@takii.kmu.ac.jp, Tel: 06-992-1001, Fax: 06-992-1781)

FEATURES Location/Qualifiers

source 1..1209

/organism="Bos taurus"

/db_xref="taxon:9913"

210..740

/codon_start=1

/product="nociceptin/orphanin FQ precursor"

/db_xref="PID:d1026347"

/db_xref="PID:g3041769"

/translation="MKILFCDDVLLSLFSSVSSCORDCITCOEKLHPADSPLEVC ILCEEKVPPSPPLMTPTCTRYMAGSSGSLSPADPELYSALIQKASEMOKLRMPYR SLVQVDAEPGADAEPADEPADEAEVQKQKQKRGFTGARSARKLANQKRFSEFMROYLVLS MSSQRRLHONGNV"

BASE COUNT 240 a 359 c 341 g 269 t

ORIGIN

```

Query Match          92.6%; Score 199; DB 20; Length 1209;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 654 TTCAGTACGATTATGAGCAGTACCGGCTCCGAGCATGCAAGTCCAGCAG 704
      1 TTYWSMGATTTATGMCNCARTATYTTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 12
LOCUS      MMPNEX3      1253 bp      DNA      ROD      10-SEP-1996
DEFINITION M.musculus gene encoding prepronociceptin, exon 3.
ACCESSION  X97372
NID         91532038
KEYWORDS   prepronociceptin.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 1253)
AUTHORS   Molleau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
            Meunier,J.C. and Parmentier,M.
            Structure, tissue distribution, and chromosomal localization of the
            prepronociceptin gene
            Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
JOURNAL    96323281
MEDLINE
REFERENCE 2 (bases 1 to 1253)
AUTHORS   Parmentier,M.
            Direct Submission
            Submitted (19-APR-1996) M. Parmentier, Universite Libre de
            Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-
            1070 Bruxelles, BELGIUM
JOURNAL
FEATURES
    source      Location/Qualifiers
                1..1253
                /organism="Mus musculus"
                /strain="129/Sv"
                /db_xref="taxon:10090"
                /clone_lib="lambda FIXII (Stratagene #946305)"
                /clone="MG43"
                605..1087
                /number=3
                /label=ex3
                /usedin=X97371:MMPN.CDS
                /usedin=X97371:MMPN.mRNA
                /product="prepronociceptin"

BASE COUNT    356 a      307 c      323 g      267 t

Query Match          92.6%; Score 199; DB 29; Length 1253;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 956 TTCAGTACGATTATGAGCAGTACCGGCTCCGAGCATGCAAGTCCAA 1006
      1 TTYWSMGATTTATGMCNCARTATYTTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 13
LOCUS      MUSHOP      1354 bp      mRNA      ROD      09-APR-1997
DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
ACCESSION  D82866
NID         91311472
KEYWORDS   nociceptin/orphanin FQ.
SOURCE     Mus musculus adult brain cDNA to mRNA.
ORGANISM   Mus musculus
            Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
            Murinae; Mus.
            1 (bases 1 to 1354)
            Direct Submission
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES
    source      Location/Qualifiers
                1..1354
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /dev_stage="adult"
                /tissue_type="brain"
                360..923
                /codon_start=1
                /product="nociceptin/orphanin FQ"
                /db_xref="PID:d1012281"
                /db_xref="PID:g1311473"
                /translation="MKIRFCDVLLSLSSVSSCCPDCITCOEKLHPAPDSFNKTC
                ILCEERKVPRLPIVYCTKYMASGSGSLSPADELYSALYOKASEMOLKRPVVR
                SLVQVRAEFGADPEADAEVEYOKLOKRFGGFGARSKARKLANQRF
                SEFRKQYLVKMSQSQRRLHONGNV"
                1317..1322

BASE COUNT    309 a      370 c      386 g      289 t
ORIGIN

Query Match          92.6%; Score 199; DB 29; Length 1354;
Best Local Similarity 49.0%; Pred. No. 3,66e-07;
Matches 25; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Db 837 TTCAGTACGATTATGAGCAGTACCGGCTCCGAGCATGCAAGTCAA 887
      1 TTYWSMGATTTATGMCNCARTATYTTNGTNTYWSNATGCAWSMNSNCAR 51

RESULT 14
LOCUS      HS1170K4      294095 bp      DNA      HTG      03-DEC-1998
DEFINITION Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
ACCESSION  AL022314
NID         93947782
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 294095)
            Direct Submission
            Submitted (03-DEC-1998) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On Dec 2, 1998 this sequence version replaced g1:3927921.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known: 800 nt's separate
            segments. Unfinished: dj1170K4 Contig_ID: 03636 acc-AL022314
            Length: 20 bp unfinished: dj1170K4 Contig_ID: 02957 acc-AL022314
            Length: 17 bp unfinished: dj1170K4 Contig_ID: 00676 acc-AL022314
            Length: 19 bp unfinished: dj1170K4 Contig_ID: 03380 acc-AL022314
            Length: 19 bp unfinished: dj1170K4 Contig_ID: 04758 acc-AL022314
            Length: 22 bp unfinished: dj1170K4 Contig_ID: 02609 acc-AL022314
            Length: 22 bp unfinished: dj1170K4 Contig_ID: 03488 acc-AL022314
COMMENT
AUTHORS
TITLE
JOURNAL

```

Length: 20 bp Unfinished: dj1170K4 Contig_ID: 02907 acc-AL022314
 Length: 24 bp Unfinished: dj1170K4 Contig_ID: 01822 acc-AL022314
 Length: 19 bp Unfinished: dj1170K4 Contig_ID: 02655 acc-AL022314
 Length: 24 bp Unfinished: dj1170K4 Contig_ID: 04353 acc-AL022314
 Length: 29 bp Unfinished: dj1170K4 Contig_ID: 03931 acc-AL022314
 Length: 20 bp Unfinished: dj1170K4 Contig_ID: 01036 acc-AL022314
 Length: 21 bp Unfinished: dj1170K4 Contig_ID: 00860 acc-AL022314
 Length: 28 bp Unfinished: dj1170K4 Contig_ID: 02949 acc-AL022314
 Length: 20 bp Unfinished: dj1170K4 Contig_ID: 00341 acc-AL022314
 Length: 42 bp Unfinished: dj1170K4 Contig_ID: 03201 acc-AL022314
 Length: 33 bp Unfinished: dj1170K4 Contig_ID: 04365 acc-AL022314
 Length: 17 bp Unfinished: dj1170K4 Contig_ID: 05023 acc-AL022314
 Length: 3767 bp Unfinished: dj1170K4 Contig_ID: 01947
 acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 01270
 acc-AL022314 Length: 19 bp Unfinished: dj1170K4 Contig_ID: 01583
 acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 03812
 acc-AL022314 Length: 29 bp Unfinished: dj1170K4 Contig_ID: 04647
 acc-AL022314 Length: 26 bp Unfinished: dj1170K4 Contig_ID: 02527
 acc-AL022314 Length: 25 bp Unfinished: dj1170K4 Contig_ID: 03783
 acc-AL022314 Length: 20 bp Unfinished: dj1170K4 Contig_ID: 01804
 acc-AL022314 Length: 20 bp Unfinished: dj1170K4 Contig_ID: 04857
 acc-AL022314 Length: 33 bp Unfinished: dj1170K4 Contig_ID: 03559
 acc-AL022314 Length: 18 bp Unfinished: dj1170K4 Contig_ID: 00485
 acc-AL022314 Length: 19 bp Unfinished: dj1170K4 Contig_ID: 02367
 acc-AL022314 Length: 25 bp Unfinished: dj1170K4 Contig_ID: 02270
 acc-AL022314 Length: 23 bp Unfinished: dj1170K4 Contig_ID: 00018

*** WARNING: Phase 1 High Throughput Genome Sequence ***

*** This sequence is unfinished. When sequencing is complete, * the sequence data presented in this record will be replaced * by a single finished sequence with the same accession number.

FEATURES
 source
 1. .294095
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="1170K4"

BASE COUNT 64545 a 71590 c 69905 g 62435 t 25620 others

ORIGIN

Query Match

Best Local Similarity 57.7%; Score 124; DB 18; Length 294095;

Matches 18; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

Db 272427 CTCAGGCTCATAGCGCATACCTGCGCTGATGATGACAGATGACA 272474

Oy 3 YWSNCAATYATGNCNCARATAYTNGTNTYTNMSNATGCAFMNSMNSCA 50

RESULT 15

LOCUS AC002541 166983 bp DNA PRI 12-SEP-1997
 DEFINITION Human BAC clone RG043K06 from 7q21-q22, complete sequence.
 AC002541
 ACCESION 92393735

KEYWORDS
 Htg.

SOURCE
 human.
 Homo sapiens

ORGANISM
 Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
 1 (bases 1 to 166983)
 Minx, P., Tin, A., Sutterer, C. and Hawkins, M.

JOURNAL
 The sequence of H. sapiens BAC clone RG043K06
 Unpublished (1997)

REFERENCE
 2 (bases 1 to 166983)
 Waterston, R.

JOURNAL
 Direct Submission
 Submitted (12-SEP-1997) Department of Genetics, Washington
 University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT
 SUBMITTED BY:
 Genome Sequencing Center
 Department of Genetics

Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:sapiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 88:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RG043K06; actual end is at 166983 of RG043K06. The orientation of this clone is unknown.

FEATURES
 This clone contains STS SW551326 (NTD:940334).

source
 Location/Qualifiers
 1. 166983

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"

/clone="RG043K06"
 /clone_1bp="CTTB-978SK-B"
 /map="7q21-q22"

repeat_region
 650..703

repeat_region
 /rpt_family="L1"
 1026..1751

repeat_region
 /rpt_family="L1"
 2172..5929

repeat_region
 /rpt_family="L1"
 complement(4862..5266)
 /rpt_family="L1"

repeat_region
 6072..6164

repeat_region
 /rpt_family="L1"
 6304..6584

repeat_region
 /rpt_family="L1"
 7600..7637

repeat_region
 /rpt_family="L1"
 11880..11901

repeat_region
 /rpt_family="L1"
 11978..12442

repeat_region
 /rpt_family="L1"
 12523..12812

repeat_region
 /rpt_family="ALU"
 complement(20178..20466)

repeat_region
 /rpt_family="ALU"
 23366..24826

```

repeat_region      /rpt_family="L1"
26530. .26570
/rpt_family="L1"
repeat_region      complement(30991. .31369)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
30991. .31392
/rpt_family="L1"
repeat_region      /rpt_family="L1"
34123. .34162
/rpt_family="L1"
repeat_region      complement(34281. .34405)
/rpt_family="L1"
repeat_region      complement(35591. .35645)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
38226. .38311
/rpt_family="L1"
repeat_region      /rpt_family="L1"
39072. .39461
/rpt_family="L1"
repeat_region      complement(40536. .41041)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
43083. .43590
/rpt_family="L1"
repeat_region      /rpt_family="L1"
44156. .44186
/rpt_family="L1"
repeat_region      complement(44592. .44618)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
45084. .45112
/rpt_family="L1"
repeat_region      /rpt_family="L1"
47602. .47635
/rpt_family="L1"
repeat_region      /rpt_family="L1"
49901. .50192
/rpt_family="L1"
repeat_region      /rpt_family="ALU"
50767. .51049
/rpt_family="ALU"
repeat_region      complement(53778. .53816)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
55314. .55367
/rpt_family="L1"
repeat_region      /rpt_family="L1"
55997. .57239
/rpt_family="L1"
repeat_region      complement(57235. .57272)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
57271. .57305
/rpt_family="L1"
repeat_region      /rpt_family="L1"
57853. .58150
/rpt_family="L1"
repeat_region      /rpt_family="ALU"
58899. .58930
/rpt_family="L1"
repeat_region      complement(58936. .59015)
/rpt_family="L1"
repeat_region      complement(59214. .59271)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
60343. .60364
/rpt_family="L1"
repeat_region      complement(61043. .61247)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
61043. .61248
/rpt_family="L1"
repeat_region      /rpt_family="L1"
64833. .64870
/rpt_family="L1"
repeat_region      /rpt_family="L1"
66745. .66929
/rpt_family="L1"
repeat_region      /rpt_family="L1"
68562. .68595
/rpt_family="L1"
repeat_region      /rpt_family="L1"
68562. .68595
/rpt_family="L1"
repeat_region      complement(68602. .68886)
/rpt_family="ALU"
repeat_region      /rpt_family="L1"
69060. .69201
/rpt_family="L1"
repeat_region      complement(69202. .69416)
/rpt_family="ALU"
repeat_region      /rpt_family="L1"
69418. .70100
/rpt_family="L1"
repeat_region      complement(69534. .69940)
/rpt_family="L1"

```

```

repeat_region      70118. .70278
/rpt_family="L1"
repeat_region      complement(70239. .70749)
/rpt_family="L1"
repeat_region      /rpt_family="L1"
70320. .71433
/rpt_family="L1"
repeat_region      complement(71540. .71685)
/rpt_family="MER"
repeat_region      complement(71880. .72079)
/rpt_family="MER"
repeat_region      complement(72426. .72561)
/rpt_family="MER"
repeat_region      /rpt_family="MER"
72472. .72930
/rpt_family="MER"
repeat_region      /rpt_family="MER"
73073. .73174
/rpt_family="MER"
repeat_region      /rpt_family="MER"
73179. .73468
/rpt_family="ALU"
repeat_region      /rpt_family="MER"
73488. .73561
/rpt_family="MER"
repeat_region      /rpt_family="MER"
73562. .73605
/rpt_family="L1"
repeat_region      /rpt_family="L1"
73606. .73766
/rpt_family="MER"
repeat_region      /rpt_family="L1"
74103. .74358
/rpt_family="L1"
repeat_region      /rpt_family="L1"
74492. .75185
/rpt_family="L1"
repeat_region      complement(77916. .78147)
/rpt_family="ALU"
repeat_region      /rpt_family="L1"
78284. .78313
/rpt_family="L1"
repeat_region      complement(79753. .79791)

```

Note: remainder of annotations omitted.

```

Query Match      55.3%; Score 119; DB 27; Length 166983;
Best Local Similarity 36.7%; Pred. No. 1.83e+01;
Matches 18; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

DB 79936 GGCATCATGACTTTTAAAGCCCAATAGTAAAGCATAAACCTGAANA 79984
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
CP 49 GNSMNSWITGCATNSWMAACNARITATGNCCKATRAATGNSWRAA 1

```

Search completed: Tue Apr 27 10:23:42 1999
Job time : 146 secs.

 WIRENET (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:26:44 1999; MasPar time 25.91 Seconds
 267.841 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 215
 N.A. Sequence: 1 TTWVNSGARTTATGNGNCA.....TWSNATGCAKSNMSNCAR 51
 Comp: AARWNSCTYAARTACKNGT.....AMWSNACTGYTWSNNGRY

Scoring table: TABLE bkttranslated2
 Gap 40

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: n-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 37.150; Variance 152.162; scale 0.244

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
C	1	199	92.6	932	34	T79888		DNA encoding human OP	3.11e-07
C	2	108	50.2	11517	4	Q26021		PSP6-SFV4 RNA transcr	1.18e+01
C	3	106	49.3	1873	13	Q80220		Human NDF-beta1a cion	1.67e+01
C	4	105	48.8	988	31	T43486		ARM gene exons 53-54.	1.98e+01
C	5	104	48.4	271	2	M60950		Derivative of plasmid	2.34e+01
C	6	103	47.9	3016	19	T17115		Rhodopsin gene.	2.78e+01
C	7	102	47.4	2504	15	Q90443		Murine oviduct specif	3.29e+01
C	8	99	46.0	405	10	Q44373		Sequence of VHN DNA s	4.45e+01
C	9	98	45.6	315	8	Q60950		Human brain Expressed	6.44e+01
C	10	98	45.6	1692	1	Q03702		Gene encoding plant e	6.44e+01
C	11	98	45.6	2301	22	T32324		Soluble starch syntha	6.44e+01
C	12	98	45.6	3314	10	Q58896		Humanised anti-CEA SF	6.44e+01

13	98	45.6	3540	2	Q11854	Glutamate receptor 6.	6.44e+01
14	98	45.6	4127	31	T68646	Potato soluble starch	6.44e+01
15	97	45.1	423	33	T80701	Type II topoisomerase	7.60e+01
16	97	45.1	540	11	Q62072	L. heterochlochi JCM11	7.60e+01
17	97	45.1	540	15	Q88205	Lactobacillus sp. 16S	7.60e+01
18	97	45.1	1133	36	T74986	Human H1A DR alpha te	7.60e+01
19	97	45.1	1972	8	Q49050	Fungus-responsive pot	7.60e+01
20	97	45.1	5737	17	T12235	pp60Pik cDNA.	7.60e+01
21	96	44.7	2735	2	Q11072	B.steatothermophilus	8.97e+01
22	95	44.2	356	6	Q39131	Mouse MK coding seque	1.06e+02
23	95	44.2	440	36	T88799	Nuclear steroid hormo	1.06e+02
24	94	43.7	1365	1	Q04715	Human arginase cDNA c	1.24e+02
25	94	43.7	1407	15	Q92081	Human pexoxisome proI	1.24e+02
26	94	43.7	3614	35	T91883	Murine SIM (single-mi	1.24e+02
27	94	43.7	3981	1	N70695	Macrophage-colony sti	1.24e+02
28	94	43.7	3981	1	N91734	Sequence of a macroph	1.24e+02
29	94	43.7	4021	2	Q10343	Monocyte macrophage c	1.24e+02
30	94	43.7	7454	20	T29660	Nisin nisAaNCIP gene	1.24e+02
31	93	43.3	274	34	T61624	Streptococcus thermop	1.46e+02
32	93	43.3	403	8	Q60838	Human brain Expressed	1.46e+02
33	93	43.3	413	40	V16480	DNA encoding a Bac11	1.46e+02
34	93	43.3	452	20	T22420	Human gene signature	1.46e+02
35	93	43.3	652	5	Q34564	Shrgp.	1.46e+02
36	93	43.3	1313	27	T48101	Immunogenic type F bo	1.46e+02
37	93	43.3	1449	32	T72162	N-acetylglucosamine 1	1.46e+02
38	93	43.3	2799	29	T42854	Phospholipase D gene	1.46e+02
39	93	43.3	2799	16	Q08675	DNA encoding Phosphol	1.46e+02
40	93	43.3	3075	37	V06252	Murine purmycin-sens	1.46e+02
41	93	43.3	3695	6	Q32844	Encodes E. coli heat	1.46e+02
42	93	43.3	3762	6	Q38738	Encodes E. coli heat	1.46e+02
43	93	43.3	5579	23	T30056	S. pneumoniae detecti	1.46e+02
44	93	43.3	6953	19	T17116	Rhodopsin gene.	1.46e+02
45	92	42.8	8585	2	N60404	Sequence encoding hum	1.72e+02

ALIGNMENTS

RESULT 1
 ID T79888 standard; DNA; 932 BP.
 AC T79888;
 DT 08-DEC-1997 (first entry)
 DE DNA encoding human opioid receptor-like 1 receptor ligand.
 KW Opioid receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 OS Synaptic.
 FH Key Location/Qualifiers
 FT cds 1..507
 FT /tag= a
 FT /note= "no start codon"
 PN W09707208-A1.
 PD 27-FEB-1997.
 PF 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368.
 PA (ULBR) UNTV LIBRE BRUXELLES.
 PI Meunier J, Molliereau C, Parmentier M, Vassart G;
 DR WPI: 97-165292/15.
 DR P-PSDB: W25162, W25163, W25164.
 PT Novel ligand for the opioid receptor-like receptor. nociceptin -
 PT has pronociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
 PT etc.
 PS CLAIM 1: Page -? 48pp; English.
 CC T79888 encodes a ligand to human opioid receptor-like 1 (ORL1) receptor
 CC designated nociceptin, which resembles the endorphin dynorphin A.
 CC Inhibitors of nociceptin or its DNA, e.g. antisense sequences and
 CC antibodies can be used as new types of drugs in the control of various
 CC behaviours or functions. The inhibitors can be used to prevent or treat
 CC a disease related to hyperalgesia, neuroendocrine secretion, stress,
 CC locomotor activity, anxiety, instinctive behaviour, learning and memory,
 CC homeostasis, hyperalgesia, hypalgesia and/or sensory perception.
 SQ Sequence 932 BP: 224 A: 273 C: 238 G: 197 T:

RESULT	4
ID	T43486 standard; DNA; 988 BP.
AC	T43486;
DT	11-AUG-1997 (first entry)
DE	ATM gene exons 53-54.
KW	ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody
KW	central nervous system; immune system; chromosome instability; therapy
KW	cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW	multi-system disease; autosomal recessive; cerebellar ataxia; cerebellu
KW	general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia
OS	blood vessel; bulbar conjunctiva; facial skin; A-T; ss.
HF	Homo sapiens.
FT	Key
FT	intron
FT	1..234
FT	/tag= a
FT	/number= 52
FT	235..348
FT	/tag= b
FT	/number= 53
FT	349..669
FT	/tag= c
FT	/number= 53
FT	670..828
FT	/tag= d
FT	/number= 54
FT	829..988
FT	/tag= e
FT	/number= 54
PN	MO636691-A1.
PD	21-NOV-1996.
PE	16-MAY-1996; U07025
PR	16-MAY-1995; US-441822.
PR	08-APR-1996; US-629001.
PA	(KOHN/) KOHN K I.
PA	(UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI	Shlloh Y;
DR	WPI; 97-012070/01.
PT	New isolated ataxia-telangiectasia gene - used to develop prods. for
PS	the study, diagnosis and treatment of ataxia-telangiectasia.
CC	claim 1; Page 70; 13pp; English.
CC	T43444;T43496 represent exons of the ATM gene of the invention.
CC	Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC	deletions in the coding region of the ATM gene. A-T is a progressive
CC	genetic disorder affecting the central nervous and immune systems. A-T
CC	involves chromosomal instability, cancer predisposition, radiation
CC	sensitivity, and cell cycle abnormalities. A-T is a multi-system disease


```

RESULT 14
ID T68646 standard; CDNA: 4127 BP.
AC T68646;
DT 16-AUG-1997 (first entry)
DE Potato soluble starch synthase CDNA.
KW Starch synthase; transgenic plant; potato; rice; Oryza sativa;
KW tomato; Lycopersicon esculentum; wheat; Triticum aestivum; cassava;
KW Manihot esculenta; sweet potato; Ipomoea batatas; barley;
KW Hordeum vulgare; oat; Avena; maize; Zea mays; ds.
OS Solanum tuberosum cv. Desiree.
FH Key Location/Qualifiers
FT cds 143..3835
FT signal_peptide 143..322
FT /tag= a
FT /tag= b
FT mat_peptide 323..3832
FT /tag= c
PN EP-779363-A2.
PD 18-JUN-1997.
PF 11-DEC-1996; 309004.
PR 12-DEC-1995; GB-025353.
PA (NAT) NAT STARCH & CHEM INVESTMENT HOLDING COR.
PI Edwards EA, Marshall J, Martin CR, Smith AM.
DR MPI: 97-312737/29.
DR P-PSDB: W17765.
PT Soluble starch synthase - used to produce altered starch from
PT commercially important plants, e.g. potato, rice, wheat, and maize
PS Claim 16; Page 18-24; 39pp; English.
CC A composite cDNA sequence (T68646) codes for a 100-140 kDa potato
CC soluble starch synthase (W17765). The sequence was deduced from 2
CC overlapping cDNA clones obtd. from a tuber cDNA library by
CC immunoscreening with antiserum raised against purified starch
CC synthase. Starch synthase nucleic acids can be linked to a plant
CC promoter, in sense or antisense direction, and used to alter starch
CC characteristics in transgenic plants, esp. potato, tomato, rice,
CC wheat, pea, cassava, sweet potato, barley, oat or maize. A novel
CC altered starch, extracted from such transgenic plants, has a
CC viscosity onset temperature that is reduced by at least 5 degC
CC compared to starch from non-transformed plants.
SO Sequence 4127 BP; 1282 A; 709 C; 977 G; 1159 T;

Query Match 45.6%; Score 98; DB 31; Length 4127;
Best Local Similarity 37.8%; Pred. No. 6.44e+01;
Matches 17; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

Db 3576 gactttagatcgtgattgattgacatgacaagaagagag 3620
OY 7 GARTTATGMCNARTATYNTGNTWMSNATGCARWSNMCAR 51

RESULT 15
DT 780701 standard; DNA: 423 BP.
AC 780701;
DT 11-NOV-1997 (first entry)
DE Type II topoisomerase database reference sequence SEO ID NO:120.
KW Detection; identification; TopoII; contamination; food; human;
KW animal; plant; soil; water; PCR; polymerase chain reaction; ds.
OS Borrelia garinii.
PN US5645994-A.
PD 08-JUL-1997.
PF 05-JUL-1990; 548138.
PR 06-JUN-1995; US-470179.
PR 05-JUL-1990; US-548138.
PR 13-AUG-1993; US-106482.
PA (UTAH) UNIV UTAH RES FOUND.
PI Huang WM.
DR MPI: 97-362925/33.
PT Detection and identification of organisms - using DNA primers to
PT amplify signature segment of organism's type II topoisomerase
PS Claim 6; Column 121-124; 114pp; English.
CC A method has been produced for selectively amplifying DNA segments of

```

```

CC one or more species of organisms in a sample. The method involves: (a)
CC providing a database containing reference sequences, comprising a
CC subunit sequence of a signature region of a macromolecule selected from
CC a type II topoisomerase (TopoII), or a homologue, where each reference
CC sequence is specific to a different species of a chosen group, and the
CC macromolecule comprises 1st and 2nd conserved regions adjacently
CC flanking the signature region; and (b) making an extract of DNA
CC molecules, and selectively amplifying DNA segments of the signature
CC region using a universal primer composition, comprising a primer
CC constructed to bind a DNA encoding the macromolecule, to produce
CC amplified DNA segments. The present sequence represents a DNA fragment
CC containing a signature segment which is used in the database as a
CC reference sequence for Borrelia garinii. The method can be used to
CC identify all of the different organisms present in a single
CC sample without using multiple probes. It can accurately distinguish
CC among similar and related species, and can be used with, e.g.
CC contaminated food products, tissue or fluid samples from diseased
CC humans, animal or plants, soil samples and water samples from any
CC source.
SO Sequence 423 BP; 125 A; 64 C; 94 G; 140 T;

Query Match 45.18; Score 97; DB 33; Length 423;
Best Local Similarity 41.7%; Pred. No. 7.60e+01;
Matches 15; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

Db 16 gtccatagagagaacttattctctatgtatgaatg 51
OY 9 RTTATGMCNARTATYNTGNTWMSNATGCARWS 44

Search completed: Tue Apr 27 10:27:17 1999
Job time : 33 secs.

```


Contact: Marra M/Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseat@watson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 203590

Seq primer: mob.REGA+ET
High quality sequence stop: 382
Location/Qualifiers
1..388

/organism="Mus musculus"
 /note="Vector: pT73 (Pharmacia) with a modified
 polylinker: Site 1: Not I; Site 2: Eco RI; left strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGATGAGGAGGCGCCGCAATTTTCTTTTCTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."
 /date="19960601"

```

100 a      98 c      96 g      94 t
      /clone_1lb="Soares mouse p3NMFl9.5"
      /dev_stage="19.5 dpc total fetus"
      /lab_host="DH10B (ampicillin resistant)"

```

Query Match	52.6%;	Score 113;	DB 10;	Length 388
Best Local Similarity	42.9%;	Pred. No. 3.28e-03;		
Matches	18;	Conservative	13;	Mismatches 13;

Dd 6 TGAAGATACATCCTGTGTACCAAGAAATGTCACATGAAC 4
 || : :: ||| :: : || : ||| | ||| ::
Cp 50 TGNMNSWYTCATNSWNAARNACNARATAYGNCACATRAY 9

RESULT	5				
LOCUS	AA796248	421 bp	mRNA	EST	09-FEB-1998
DEFINITION	v998c09.i1 Barstead mouse myotubes MPRB5 Mus musculus cDNA clone				
DEFINITION	1154370 5', mRNA sequence.				
ACCESSION	AF006249				

NID 92859203
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 421)

TITLE	COMMENT
The WashU-HHMI Mouse EST Project Unpublished (1996)	

Contact: Maria M/Mouse EST Project
WASHU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNM; con

Source

IMAGE Consortium (image.llnl.gov) for further information
 MGI:627528
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 419
[†]<http://www.ncbi.nlm.nih.gov/seq/trace/627528>

1. .421
/organt

```

/organism="Mus musculus"
/strain="C3H"
/notes="Vector: pT73D-PAC (Pharmacia) with a modified
polylinker. Site_1: EcoRI. Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACGATCTGAGAGGAGGAGGCGCCGCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(PAATCGATCGTC), digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Bartsed. The C2C12 cell line
available from ATCC, catalog # CRL-1772) differentiates
rapidly, forming contractile myotubes and producing
characteristic muscle proteins."
/db_xref="taxon:10090"

```

BASE COUNT	103 a	108 c	106 g	104 t
ORIGIN				

Query Match	52.68;	Score 113;	DB 12;	Length 421;
Best Local Similarity	42.98;	Pred. No. 3.28e-03;		
Matches	18;	Conservative	12;	Mismatches 12;
				Indels 0;
				Gaps 0

Db 52 TGAAGATACCATCCTGTTCACAGAATTTGCATGTAAC 9
 || : :: ||| : : | : | : ||| : :
Cp 50 TGNSWWSWTGCATNSWNARNACNAPRTAATGNCXCATPAAY 9

LOCUS	6	AA796116	470 bp	mRNA	EST	09-FEB-1998
DEFINITION		v066e06.1	Soares mouse mammary gland	NCIMG Mus musculus	CDNA clone	
ACCESSION		1054882.5	mRNA sequence.			
		AA796116				

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE
AUTHORS

Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 470)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuq

TITLE	JOURNAL	COMMENT
The WashU-HHMI Mouse Est Project	Unpublished (1996)	

Contact: Maira M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:586458
Seq primer: -28ml3 rev2 ET from Amersham.
Location/Oval:tfare

```

FEATURES
SOURCE
    Location/Qualifiers
        1..470
            /organism="Mus musculus"
            /strain="C57BL/6j"
            /note="Organ: mammary gland; Vector: pT73D-Pac
            (Pharmacia) with a modified polylinker; Site_1: Not I

```

```

Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer 15',
TGTTCGCAATCTGAGCTGAGAGCGGCCCGCATGCTTTTCTTTTCTTTTCTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pTV73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. library
constructed and normalized by Bento Soares and M. Fatma
Bonaldo.
/db_xref="taxon:10090"
/clone="1054882"
/clone_lib="Soares mouse mammary gland NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"

```

BASE COUNT

115 a 116 c 115 g 124 t

Query Match	52.6%	Score 113	DB 12	Length 470;
Best Local Similarity	42.9%	Pred. No. 3,286-03;		
Matches	18;	Conservative	12;	Mismatches 12;
			Indels	0;
			Gaps	0;
Db	87	TGGAAGCATGCACTCTGTGTACAGCAATTTGTCAATGAAC	128	
		::::: :::	:::	
Cp	50	TGMSWMSWTCATGNSMNRACNNARATYTCNCCARPAAY	9	

RESULT	7	AA269345	473 bp	MRNA	EST	26-MAR-1997
LOCUS		vb09g05.r1	Soares mouse	MMU Mus musculus	CDNA	clone 748472 5', MRNA
DEFINITION						
ACCESSION		AA269345				
NID		q1907624				
KEYWORDS		EST.				
SOURCE		house mouse.				
ORGANISM		Mus musculus				

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (Pages 1 to 473)
Marrir, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Knudab, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theusling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The Wasnu-RHMI Mouse Est Project
Unpublished (1996)

Contact: Maria M/Mouse EST Project
 MashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:457456
 Seq primer: -28ml3 rev2 ET from Amersham
 high quality sequence stop: 466.
 Location/Qualifiers
 1..473
 source

```

/organism="Mus musculus"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAAGGCGCGCGCCAACTTTTTTTTTTTTTTTT
3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

```

```

      /db_xref="taxon:10090"
      /clone_id="Scars mouse NML"
      /tissue_type="Liver"
      /lab_host="DH10B"
      <1..>473
      mRNA
      BASE COUNT      110 a      106 c      115 g      142 t
      ORIGIN

```

	Query Match	52.6%	Score 113	DB 24	Length 473
	Best Local Similarity	42.9%	Pred. No.	3	28
	Matches	18	Conservative	12	Mismatches 12
				Indels	0
					Gaps 0
Db	402	TGGAAGGATAGCATCCTGTGTACACAAGAAATGTACATGAAC	443		
	: : : : : : : : : : : : :				
Cp	50	TGNSMNSWYTCATCATNSWARRNACARRAATYAGNCKCATFRAY	9		

RESULT	8						
LOCUS	AAT95004	488 bp	mRNA	EST	09-FEB-1998		
DEFINITION	VS08607.t1 Barstead mouse irradiated colon MFLR57 Mus musculus cDNA						
ACCESSION	U08607.1, mRNA sequence.						
NID	AAT95004						
KEYWORDS	G2857955						
EST.							
SOURCE	house mouse.						
ORGANISM	Mus musculus						

REFERENCE	AUTHORS
1 (pages 1 to 488)	Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kuwaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennou, G., Soares, B., Wilson, R. and Waterston, R.
TITLE	The Washou-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watsn.wustl.edu
 This clone is available royalty-free through LILN; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:618684
 Putative full length read
 vector to vector length is 564
 Seq primer: -28m3 rev2 ET from Amersham
 High quality sequence stop: 463.
 Location/Qualifiers
 1..488

```

/organism="Mus musculus"
/strain="FVB/N"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI. Tissue obtained
from 8 week old mouse. Colon was harvested 72 hours after
irradiation with 1400 Gys. 1st strand cDNA was primed
with a Not I - oligo(dN) primer
15'ggtttacgaactctgaaagtgagcgcgcgcgcctttttttttttttttttttt
t 3'); double-stranded cDNA was ligated to Eco RI
adaptors (taattcgatctctgg), digested with Not I and
into the Not I and Eco RI sites of the modified pT73
vector. Library constructed by Bob Barstead."
/db_xref="taxon:10090"
/clone="1137612"
/clone_11b="Barstead mouse irradiated colon MFLB7"
/dev_stage="8 weeks"
/lab_host="DH10B"

```

Query Match	52.6%	Score 113	DB 12	Length 488
Best Local Similarity	42.9%	Pred. No. 3.28e-03		
Matches	18	Conservative	12	Mismatches 12; Indels 0; Gaps 0
Db	58	TGNAAGATAGCATCCGTGTACACAAATATGTCATGACATGAC	99	
Cp	50	TGNSMNSWYTGCAATNSMNNACNNARATATGTCACATTAAT	9	
RESULT	9			
LOCUS	M62849	426 bp	MRNA	EST
DEFINITION	Y283B06.s1 Homo sapiens cDNA clone 289619.3			01-MAR-1996
ACCESSION	M62849			
NID	g1210678			
KEYWORDS	EST.			
SOURCE	<p>human clone-289619 primer-m13 -40 forward library-Soares multiple sclerosis 2NBHSP vector-PTT3P (Pharmacia) with a modified polylinker V.TYPE: phagemid host-DH10B (ampicillin resistant) Bst1-Not I Rse12-Eco RI 46 year old male. 1st strand cDNA was primed with a Not I -Oligo(dT) primer (5'-TGTTACCAATGTGAGTGGAGCGCCGACATTTTTTTTTTTTTTTT-3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and EcoRI sites of a modified pT7T vector (Pharmacia). Library went through one round of normalization to a Cot -5. Library constructed by Bento Soares and M.Felima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH).</p>			
ORGANISM	<p>Homo sapiens Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostoma; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Channata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 426) Hillier, L., Clark, N., Dubaque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Ritzkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisks, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R. The WashU-Merck EST Project Unpublished (1995)</p>			
TITLE	Contact: Wilson RK			
JOURNAL	WashU-Merck EST Project			
COMMENT	<p>Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu High quality sequence stops: 380 Source: IMAGE Consortium, LINTL This clone is available royalty-free through LINTL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Location/Qualifiers 1. 426 /organism="Homo sapiens" /clone="289619" <1..>426</p>			
FEATURES	<p>source 1. 426 /organism="Homo sapiens" /clone="289619" <1..>426</p>			
BASE COUNT	150 a	96 c	76 g	104 t
ORIGIN				
Query Match	52.1%	Score 112	DB 19	Length 426
Best Local Similarity	44.7%	Pred. No. 5.21e-03		
Matches	17	Conservative	11	Mismatches 10; Indels 0; Gaps 0
Db	234	ATACAAAAAATATGCTCATTAATATGACAGAA	271	
Cp	38	ATNSMNNACNNARATATGTCACATTAATGTCNSWMAA	1	
RESULT	10			
LOCUS	R37851	461 bp	MRNA	EST
				04-MAY-1995

DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE
Y698D02.s1 Homo sapiens cDNA clone 24106 3' similar to contains Alu repetitive element.	R37851	9795307	EST.	
human clone-24106 library-Soares infant brain IN1B vector-Lafmid BA host-Ph10B (ampicillin resistant) primer--21m3 Rstcl-Not I Rstcl-Hind III Whole brain from a 73 days post natal female. 1st strand cDNA was primed with a Not I - Oligo(dT) primer (5' AACTGCAAAATATCGCGCCGACGAAATTTTATTTTATTTT 3'); double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.				
ORGANISM			Homo sapiens	
REFERENCE			Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS			1 (bases 1 to 461) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucab,T., Le,M., Lennon,G., Marra,M., Persons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaski,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.	
TITLE			The Wash-Merck EST Project	
JOURNAL			Unpublished (1995)	
COMMENT			GDB: G00-396-453 Contact: Wilson RK Wash-Merck EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 347 Source: IMAGE Consortium, LNU This clone is available royalty-free through LNU ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Location/Qualifiers 1. 461	
FEATURES			source	
BASE COUNT			101 a 97 c 89 g 165 t	
ORIGIN			1. 461 /organism="Homo sapiens" /clone="24106"	
Query Match			52.1%; Score 112; DB 5; Length 461;	
Best Local Similarity			42.9%; Pred. No. 5,21e-03;	
Matches			15; Conservative 12; Mismatches 8; Indels 0; Gaps 0;	
Db			16 TTACGTATTTTTCAGACAGTATTTTCCTGTC 50	
Qy			1 TTYMSNGARTTYATGMCARTATYTTNCTNTTMS 35	
RESULT 11				
LOCUS			A0173569 541 bp DNA GSS 09-SEP-1998	
DEFINITION			HS-3202_A2.C03_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3202 Col-6 Row-E, genomic survey sequence.	
ACCESSION			A0173569	
NID			93570936	
KEYWORDS			GSS.	
SOURCE			human.	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS			1 (bases 1 to 541) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J. and Shaker,R.	
TITLE			Construction of a Characterized Clone Resource for Genomic Sequencing	
JOURNAL			Unpublished (1998)	

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618

High quality sequence stop: 541.
Location/Qualifiers
1. .541

Accession	Sex	Age	Height	Weight	Others
139 a	178 c	69 g	152 t	3	others

Query Match	52.1%;	Score 112;	DB 28;	Length 541;
Best Local Similarity	47.5%;	Pred. No. 5.21e-03;		
Matches	19;	Conservative	10;	Mismatches 11;
				Indels 0;
				Gaps 0;

9 TGCATTTGAACCCCTAGATACTCTCTCATCTACTCTCTAA 48
 ||||| : | | : : ||| | : ||| : : ||
 41 TGCATNSWNAARNACNARRTAYTGNCCKATRAAYTCNSWRA 2

RESULT	12
LOCUS	A1061964 500 bp mRNA EST
DEFINITION	LD35072_3prime ID Drosophila melanogaster embryo pot2 Drosophila
ACCESSION	melanogaster cDNA clone LD35072 3prime, mRNA sequence.
NID	A1061964
KEYWORDS	g3337803
SOURCE	EST,
ORGANISM	fruit fly, Drosophila melanogaster

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 500)	Harvey, D., Hong, L., Brooksstein, P., Lewis, S. and Rubin, G.M.	BDGP/HMMI Drosophila EST Project	Unpublished (1997)	

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://fruitfly.berkeley.edu/EST_est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or
more T residues at the beginning of the sequence, this clone
probably contains an inverted insert. The sequence has been trimmed
and the T residues removed.
Plate: 350 row: F column: 12
High quality sequence stop: 437.
Location/Qualifiers
1..500

```

/organism="Drosophila melanogaster"
/note="Organ: embryo; Vector: pOT2; Site.1: EcoRI; Site.2
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
/dl_xref="taxon:7227"
/clone="LD35072"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"

```

BASE COUNT	ORIGIN	170 a	76 c	88 g	166 t
/dev_stage="0 to 24 hours mixed stage embryonic					
/lab_host="XLI Blue"					

Query Match	51.6%	Score 11;	DB 15;	Length 500;
Best Local Similarity	46.2%;	Pred. No. 8.24e-03;		
Matches	18;	Conservative	10;	Mismatches 11;
				Indels 0;
				Gaps 0

Db 207 TGCATTCAATACAAATATTTAGTCATTAATGCTGAA 24
||||| - : | : : | : | : ||| : | : :
Cp 41 TGCATNSWNAARNACNARRTAYTGNCCKATRAAYTNSWR 3

RESULT	13			
LOCUS	AA978826	757 bp	mRNA	EST
DEFINITION	LD33002.3prime LD Drosophila melanogaster embryo pot2 Drosophila			
ACCESSION	melanogaster cDNA clone LD3302 3prime, mRNA sequence.			
NID	AA978826			
KEYWORDS	EST.			
SOURCE	Fruit fly.			
ORGANISM	Drosophila melanogaster			
	Drosophila			

REFERENCE AUTHORS	TITLE JOURNAL CONVENTION
1 (bases 1 to 757) Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G. M.	BDSF/HHMI Drosophila EST Project unpublished (1997)

Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 LSA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: fruttlify.berkeley.edu/EST_est@fruttlify.berkeley.edu
 Based upon the presence of a XhoI site followed by a run of 14 or
 more T residues at the beginning of the sequence, this clone
 probably contains an inverted insert. The sequence has been trimmed
 and the T residues removed.
 Plate: 330 row: A column: 2
 High quality sequence stop: 667.
 Location/Qualifiers
 source
 1..757

	/organism="Drosophila melanogaster"	
	/note=Organ: embryo; Vector: pOT2; Site:1: Score1: Site:2	
	Xho1; Sized fractionated cDNAs were directly ligated into	
	pOT2."	
	/db_xref="taxon:7227"	
	/clone_id="ID3302"	
	/clone_1b="LD Drosophila melanogaster embryo pOT2"	
	/sex="male and female"	
	/dev_stage="0 to 24 hours mixed stage embryonic"	
	/lab_host="Xl1 Blue"	
BASE COUNT	235 a 128 c 158 g 236 t	
ORIGIN		
Query Match	51.6%;	Score 111;
Best Local Similarity	46.2%;	DB 14;
Matches	18;	Pred. No. 8.24e-03;
	Conservative	10; Mismatches 11; Indels 0; Gaps 0.

[illegible]

THIS PAGE BLANK (USPTO)

 M O S E L L (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparc_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 16 14:25:17 1999; Maspar time 2.55 Seconds
 50.745 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.dep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 131922 segs, 16180660 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneeq2
 1:part1 2:part2 3:part3 4:part4 5:parts 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29

Statistics: Mean 14.066; Variance 35.673; scale 0.394

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	53	100.0	8 24	W25164	Human oploiid receptor	2.48e+00
2	42	79.2	551 24	W18790	Corrected Bacillus la	7.03e+01
3	41	77.4	89 12	R66032	Mammalian MEK kinase	9.41e+01
4	41	77.4	224 9	R44533	Protein required for	9.41e+01
5	38	71.7	472 24	W26312	Rat STCH chaperone pr	2.33e+02
6	38	71.7	746 7	W01225	Aminopeptidase PRAP-4	2.23e+02
7	38	71.7	740 7	R81815	Acetobacter dighanyia	2.23e+02
8	38	71.7	941 25	W31559	Xenopus frog protein	2.23e+02
9	38	71.7	1319 8	R45002	Cellulose synthase op	2.23e+02
10	37	69.8	15 16	R92307	Tryptic peptide from	2.96e+02
11	37	69.8	532 9	R47237	Wild-type Feline Hepr	2.96e+02
12	37	69.8	532 13	R86588	Feline herpesvirus g1	2.96e+02
13	37	69.8	716 13	R71616	Thermotoga pullulan	2.96e+02
14	37	69.8	816 27	W37501	Human nel-related pro	2.96e+02
15	37	69.8	839 16	R75109	Glycosyl-phosphatidyl	2.96e+02
16	37	69.8	839 4	R22275	Bovine liver GPI-PD.	2.96e+02
17	37	69.8	1129 13	R70830	Murine JAK2 kinase.	2.96e+02
18	36	67.9	243 25	W15799	Adherence factor 10AR	3.92e+02

19	36	67.9	286 1	P82590	Polypeptide with glyc	3.92e+02
20	36	67.9	309 26	W21741	Neisseria meningitidi	3.92e+02
21	36	67.9	309 13	R70762	Meningococcal group B	3.92e+02
22	36	67.9	313 26	W21744	Neisseria meningitidi	3.92e+02
23	36	67.9	316 16	R75111	Glycosyl-phosphatidyl	3.92e+02
24	36	67.9	316 4	R22277	Human pancreatic GPI-	3.92e+02
25	36	67.9	350 17	R92754	Human growth differen	3.92e+02
26	36	67.9	425 24	W27163	Human TRAF inhibitor	3.92e+02
27	36	67.9	441 26	W34856	Human tau protein.	3.92e+02
28	36	67.9	583 2	R07999	Asparagine synthetase	3.92e+02
29	36	67.9	586 2	R07998	Asparagine synthetase	3.92e+02
30	36	67.9	841 4	R22276	Human liver GPI-PD.	3.92e+02
31	36	67.9	841 16	R75110	Glycosyl-phosphatidyl	3.92e+02
32	36	67.9	1339 21	W15905	Human SKcy protein.	3.92e+02
33	36	67.9	4987 3	R10834	Ranodin receptor.	3.92e+02
34	36	67.9	5035 5	R25450	MH mutant porcine rya	3.92e+02
35	36	67.9	5072 2	R11510	Ryanodine receptor de	3.92e+02
36	35	66.0	50 26	P30660	Enkephalin-like pepti	5.16e+02
37	35	66.0	59 24	W22623	Type II topoisomerase	5.16e+02
38	35	66.0	282 17	R93450	Bacteriophage resistin	5.16e+02
39	35	66.0	351 28	W34209	Streptomyces acyltran	5.16e+02
40	35	66.0	398 4	R22322	Marek Disease Virus U	5.16e+02
41	35	66.0	512 22	W15274	Salmonella secreted p	5.16e+02
42	35	66.0	541 26	W31807	Human receptor protei	5.16e+02
43	35	66.0	833 6	R28960	Delta DII.	5.16e+02
44	35	66.0	903 28	W37391	Human additional sex	5.16e+02
45	35	66.0	903 28	W37389	Human additional sex	5.16e+02

ALIGNMENTS

RESULT 1
 ID W25164 standard; peptide: 8 AA.
 AC W25164;

DT 08-DEC-1997 (first entry)
 DE Human oploiid receptor-like 1 receptor ligand, nociceptin, fragment.
 KW Oploiid receptor-like 1; ORLI; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 KW hypoaesthesia; nociceptin; endorphin; dynorphin A.
 OS Synthetic.
 PN W09707208-A1.

PD 27-FEB-1997.
 PF 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002368.
 PA (ULBR) UNIV LIBRE BRUXELLES.
 DR Meunier J, Mollereau C, Parmentier M, Vassart G;
 DR WPI, 97-165292/15.
 DR N-PSDB; T79888.

PT Novel ligand for the oploiid receptor-like receptor, nociceptin -
 PT has pro-nociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
 PT etc.

PS Claim 8, page 27; 48pp; English.
 CC W25162-W25164 are fragments of a ligand to human oploiid receptor-like 1
 CC (ORLI) receptor, designated nociceptin, which resembles the endorphin
 CC dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense
 CC sequences and antibodies can be used as new types of drugs in the
 CC control of various behaviours or functions. The inhibitors can be used
 CC to prevent or treat a disease related to hyperalgesia, neuroendocrine
 CC secretion, stress, locomotor activity, anxiety, instinctive behaviour,
 CC learning and memory, homeostasis, hyperalgesia, hypoaesthesia and/or
 CC sensory perception.
 SO Sequence 8 AA:

Query Match 100.0%; Score 53; DB 24; Length 8;
 Best Local Similarity 100.0%; Pred. No. 2.48e+00;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tlhngnv 8
 |||||
 QY 1 TLHONGNV 8

RESULT 2
ID W18790 standard; protein; 551 AA.
AC W18790.
DT 18-NOV-1997 (first entry)
DE Corrected Bacillus laurus (NCIMB 40250) endoglucanase Endo 3A.
KW Endoglucanase; Endo 3A; formation; localised; variation;
KW colour density; surface; dye; fabric; family 5; cellulose;
KW hydrolysis; p-nitrophenyl-beta-1,4-cellobioside; stone wash;
KW blue jeans; back staining.
OS Bacillus laurus.
PN W09709410-A1.
PD 13-MAR-1997.
PE 03-SEP-1996; DK0364.
PR 08-SEP-1995; DK-000993.
PA (NOVO) NOVO-NORDISK AS.
PI Fitch M, Onishi M, Schulein M, Toft AH;
DR WPI: 97-192888/17.
PT localised variation of colour density in the surface of a dyed
PT cellulosic fabric - uses cellulase compsn. able to hydrolyse
PS p-nitrophenyl-beta-1,4-cellobioside
CC Disclosure; Pages 15-17; 23pp; English.
CC The present sequence is the corrected version of the incorrect
CC Bacillus laurus (NCIMB 40250) endoglucanase Endo 3A described in
CC W09110732. Endo 3 can be used in novel method of forming localised
CC colour density variation on the surface of a dyed cellulosic
CC fabric. The method comprises agitating the fabric in an aqueous
CC medium (pH 6.5 to 9.0) containing a family 5 cellulose,
CC e.g. the present sequence, which can hydrolyse p-nitrophenyl-beta
CC -1,4-cellobioside, or a family 7 cellulase, and a mechanical
CC abrading agent or cellulose having abrading activity. Each
CC cellulase displays 30 % or more of its maximum activity at pH 7.
CC The process is useful to provide a stone washed look to blue jeans
CC without back staining.
SQ Sequence 551 AA;

Query Match 79.2%; Score 42; DB 24; Length 551;
Best Local Similarity 75.0%; Pred. No. 7.03e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 538 lhhngntl 545
QY 1 TLHONGNV 8

RESULT 3
ID R66032 standard; protein; 89 AA.
AC R66032.
DT 28-JUN-1995 (first entry)
DE Mammalian MEK kinase (MEKK 4) catalytic kinase domain.
KW MEK kinase; MEKK 4; mitogen-activated protein kinase regulator;
KW MAPK; cell atrophy inhibition; Parkinson's; Alzheimer's; cancer;
KW autoimmune diseases; allergies; wound healing; oncogenes;
KW tumour agents; neurotropic growth factor.
OS Mus musculus.
PN W09424159-A.
PD 27-OCT-1994.
PE 15-APR-1994; U04178.
PR 15-APR-1993; US-049254.
PA (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
PI Johnson GL;
DR WPI: 94-357747/44.
PT New MEK kinase protein and related antibodies and nucleic acid
PT regulator of mitogen activated protein kinase, useful
PT therapeutically to inhibit cell atrophy, to screen for oncogenes
PT etc.
PS Claim 6; Page 15; 84pp; English.
CC 079925 encodes R66032 the mammalian MEK kinase (MEKK 1), other
CC unique mammalian MEK kinases identified by PCR are described in
CC R66030 (MEKK 2), R66031 (MEKK 3) and R66032 (MEKK 4). MEKK is an
CC activator, independent of Raf protein, of mitogen-activated protein
CC kinases (MAPK). Inactivation of MEKK can be used in the treatment
CC of some cancers, autoimmune diseases and allergies, while

CC stimulation can promote wound healing. MEKK can also be used to
CC alleviate cellular atrophy in Parkinson's or Alzheimer's by acting
CC as a neurotropic growth factor, and to screen for oncogenes and
CC tumour agents.
SQ Sequence 89 AA;

Query Match 77.4%; Score 41; DB 12; Length 89;
Best Local Similarity 71.4%; Pred. No. 9.41e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 83 lhhngnv 89
QY 2 LHONGNV 8

RESULT 4
ID R44533 standard; protein; 224 AA.
AC R44533.
DT 24-JUN-1994 (first entry)
DE Protein required for surfactin synthesis.
KW Surfactin; Bacillus subtilis; antibiotic; fungicide; anticoagulant;
KW thrombosis; myocardial infarction; embolism.
OS Bacillus subtilis.
PN EP-576050-A.
PD 29-DEC-1993.
PE 23-APR-1993; 201177.
PR 24-APR-1992; IT-MI0976.
PA (ENTE) ENRICHESPA SPA.
PI Cosmina P, De Ferra F, Grandi G, Perego M, Rodriguez F;
DR WPI: 94-001231/01.
DR N-PEDB: 039499.
PT DNA from Bacillus subtilis - encodes surfactin synthesis protein
PT Claim 1; Figure 4; 24pp; English.
CC The protein is required for surfactin synthesis. The coding
CC sequence (sfp gene, 039499) is useful for the in vitro and in vivo
CC production of surfactin and analogues having anti-cholesterol,
CC functional and antibiotic properties. It can also be used as an
CC anticoagulant for thrombosis and for prophylaxis of myocardial
CC infarcts and pulmonary emboli. This protein was isolated from a
CC producing strain of microorganism.
SQ Sequence 224 AA;

Query Match 77.4%; Score 41; DB 9; Length 224;
Best Local Similarity 71.4%; Pred. No. 9.41e+01;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 172 lhhgdqy 178
QY 2 LHONGNV 8

RESULT 5
ID W26312 standard; protein; 472 AA.
AC W26312.
DT 17-NOV-1997 (first entry)
DE Rat STCH chaperone protein.
KW STCH; stress 70 protein; chaperone protein; protein folding; rat.
OS Rattus sp.
PN US5646249-A.
PD 08-JUL-1997.
PE 28-FEB-1994; 203905.
PR 28-FEB-1994; US-203905.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Kaye FU, Otterson GA;
DR WPI: 97-362996/33.
DR N-PSDB: T84446.
PT Recombinant Stress 70 Chaperone family STCH proteins - useful as
PT chaperone proteins for facilitating protein folding
PS Claim 1; Column 47-50; 30pp; English.
CC Rat STCH (W26312) is a novel member of the stress 70 chaperone
CC protein family and is characterised by a truncated C-terminal
CC peptide binding domain, peptide-independent core ATPase activity
CC and a hydrophobic leader peptide. STCH over-expression suppresses

CC oncogene transformation. STCH expressed in cells after induction
 CC with calcium ionophore A23187 (but not following heat shock) and
 CC localises within the microsome lumen of a cell. It shows over 90%
 CC homology to human STCH (W26311). Recombinant STCH can be used to
 CC facilitate protein folding, e.g. during in vitro translation and
 CC in dilaivis following urea extraction, or for x-ray crystallography.
 CC The leader peptide can be used to express a desired protein in the
 CC endoplasmic reticulum. Antibodies raised against STCH can be used
 CC for STCH detection and cell localisation.
 SQ Sequence 472 AA;

Query Match 71.7%; Score 38; DB 24; Length 472;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 301 tlhgsagv 308
 ||||:|
 OY 1 TLHONGNV 8

RESULT 6
 ID W01225 standard; Protein; 496 AA.
 AC W01225;
 DT 16-MAR-1997 (first entry)
 DE Aminopeptidase PfAP-496.
 KW Flea; midgut; aminopeptidase; PfAP-496; vaccine;
 KM domestic animal; infestation; insecticide; protease-inhibitor;
 KM controlled release formulation; synergist.
 OS Siphonaptera sp.
 FH Key
 FT Location/Qualifiers
 FT Peptide 1..15
 FT /note- "Signal peptide"
 FT 1..151
 FT /note- "nFAP-151 (W01224, claim 71)."

PN W09611706-A1.
 PD 25-APR-1996
 PF 18-OCT-1995
 PR 18-OCT-1994; US-326773.
 PR 07-JUN-1995; US-485455.
 PR 07-JUN-1995; US-482130.
 PR 07-JUN-1995; US-485443.
 PR 07-JUN-1995; US-484211.
 PA (PARA-) PARAVAX INC.
 PA (HESK-) HESKA CORP.
 PI Arsten A, Dale B, Frank GR, Grieve RB, Heath A;
 PI Hunter SW, Rushlow KE, Stiegler GL, Yamana M;
 DR WPI: 96-221762/22.
 DR N-PSDB: T40861.

PT DNA encoding Flea serine protease and aminopeptidase - useful in
 PT vaccines to protect animals from flea infestation.
 PS Claim 71: Page 199-200; 241pp; English.
 CS This sequence (PfAP-496) represents a flea midgut aminopeptidase
 CC encoded by the nFAP-1580 gene, isolated from a flea cDNA library by
 CC homology with a cattle lens leucine-aminopeptidase conserved region.
 CC The sequence contains a signal peptide, and also contains N-terminal
 CC sequence PfAP-151 (W01224). The peptide may be used in a vaccine for
 CC protection of domestic animals from flea infestation, or in isolation
 CC of protease-inhibitors, which may be used in controlled release
 CC formulations to reduce the flea burden on and around the animal. The
 CC inhibitors may be included in insecticidal compositions to increase
 CC efficacy of other active agents, by reducing proteolytic activity in
 CC the flea midgut.

SQ Sequence 496 AA;

Query Match 71.7%; Score 38; DB 20; Length 496;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 394 tlhagrv 401
 ||||:|
 OY 1 TLHONGNV 8

RESULT 7
 ID R38155 standard; Protein; 740 AA.
 AC R38155;
 DT 13-OCT-1993 (first entry)
 DE Acetabacter diguanilate phosphodiesterase PDEA3.
 KW Cyclic diguanilate; diguanilate phosphodiesterase;
 KM diguanilate cyclase; cellulose production; cdg3 operon.
 OS Acetabacter xylinum.
 FH Key
 FT Location/Qualifiers
 FT misc_difference 119
 FT /note- "Val deduced from GT"
 FT misc_difference 713
 FT /note- "Trp deduced from GG"
 FT misc_difference 726
 FT /note- "Ala deduced from CC"

PN W09311244-A.
 PD 10-JUN-1993
 PF 14-OCT-1992; U08756.
 PR 29-NOV-1991; US-800218.
 PA (WEYE) WEYERHAEUSER CO.
 PI Ben-Bassat A, Benzman M, Calnoon RD, Gelfand DH;
 PI Tal R, Wong HC;
 DR WPI: 93-197062/24.
 DR N-PSDB: Q43562.
 FT Polynucleotide sequence from Acetabacter cdg operon - encodes
 PT cyclic di:guanosine mono:phosphate degradation enzymes e.g.
 PT 3-phosphodiesterase isozyme
 PS Claim 5: Page 86-88; 98pp; English.
 CC The amino acid sequence of protein PDEA3 was deduced from the 1st.
 CC open reading frame of the cdg3 operon. The protein is a diguanilate
 CC phosphodiesterase A, i.e. it enzymatically cleaves a single
 CC phosphodiester bond in c-di-GMP to yield the linear dimer pGpG.
 CC See also R38156.
 SQ Sequence 740 AA;

Query Match 71.7%; Score 38; DB 7; Length 740;
 Best Local Similarity 71.4%; Pred. No. 2.23e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 60 tlhagrh 66
 ||||:|
 OY 1 TLHONGNV 7

RESULT 8
 ID W31558 standard; Protein; 941 AA.
 AC W31558;
 DT 16-FEB-1998 (first entry)
 DE Xenopus frog protein "chordin"
 KW Xenopus protein; Chordin; dorsal tissue; neural tissue; vertebrate;
 KW endodermal differentiation; treatment; neurodegenerative disease;
 KW nerve cell; transforming growth factor; TGF; secreted protein.
 OS Xenopus laevis.

FH Key
 FT Location/Qualifiers
 FT Peptide 1..19
 FT /note- "Hydrophobic signal peptide"
 FT 20..941
 FT /note- "putative secreted protein chordin"

PN U55679783-A.
 PD 21-OCT-1997
 PF 22-NOV-1994; 343760.
 PR 22-NOV-1994; US-343760.
 PA (REGC) UNIV CALIFORNIA.
 PI De Robertis EM, Sasal Y;
 DR WPI: 97-525754/48.
 DR N-PSDB: T93499.
 PT DNA encoding Xenopus frog protein - that induces dorsal and neural
 PT development and endodermal differentiation in vertebrates
 PS Claim 2; Columns 23-28; 27pp; English.
 CC This is the Xenopus protein "chordin". This functional recombinant
 CC protein chordin has a defined sequence of 941 amino acids and can induce
 CC dorsal and neural development and endodermal differentiation in
 CC vertebrates. The presence of a hydrophobic signal sequence, four possible

CC N-glycosylation sites and conserved Cys-rich repeat regions suggest that
 CC chordin is a secreted protein. The encoding DNA sequence can be
 CC operationally linked with an expression vector, to form a construct and a
 CC transfectant can be obtained by introducing the construct into a host.
 CC Chordin may be useful as a component of culture media for culturing cells
 CC such as nerve or muscle cells, for treating neurodegenerative diseases
 CC and damaged nerve cells.
 SO Sequence 941 AA.

Query Match 71.7%; Score 38; DB 25; Length 941;
 Best Local Similarity 62.5%; Pred. No. 2.23e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 429 tlnhngtl 436
 QY 1 TLHONGNV 8

RESULT 9
 ID R45002 standard; Protein: 1319 AA.

AC R45002:
 DT 13-JUN-1994 (first entry)
 DE Cellulose synthase operon, gene C product.
 KW Bacterial; cellulose synthase; operon; gene A; gene B; gene C;
 KW gene D; transcription vector; recombinant microorganism;
 KW cellulose synthesis.
 OS Acetobacter xylinum.
 PN US568274-A.
 PD 07-DEC-1993.
 PF 12-APR-1989; 337194.
 PR 04-APR-1990; WO-001811.
 PR 09-APR-1990; IL-094053.
 PR 10-APR-1990; CA-014264.
 PR 11-APR-1990; IE-001317.
 PR 12-APR-1990; NZ-233312.
 PA (CERTU) CERTUS CORP.
 PI Ben-Bassat A, Ben-Ziman M, Calhoon RD, Fear AL, Gelfand DH;
 PI Meade JH, Tai R, Wong H;
 DR WPI: 93-404004/50.
 DR N-PSDB: 053522.
 PT Contiguous nucleic acid sequences - encoding bacterial cellulose
 PT synthase
 PS Claim 52; Fig 1 and Columns 53-72; 79pp; English.
 CC The sequences given in R45000-03 represent the proteins encoded by
 CC the bacterial cellulose synthase operon. The sequence given in
 CC R45004 is the beginning of an open reading frame overlapping the end
 CC of this operon. The bacterial cellulose synthase operon contains
 CC four genes, genes A-D. The operon sequence may be used in a
 CC transcription vector for the expression of the cellulose synthase
 CC operon to increase cellulose production in a recombinant microorganism.
 CC This system may be used as an important tool for exploring mechanisms
 CC of cellulose synthesis and for enhancing production of cellulose.
 SO Sequence 1319 AA;

Query Match 71.7%; Score 38; DB 8; Length 1319;
 Best Local Similarity 50.0%; Pred. No. 2.23e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 518 alqgggdv 525
 QY 1 TLHONGNV 8

RESULT 10
 ID R92307 standard; Peptide: 15 AA.

AC R92307:
 DT 17-MAY-1996 (first entry)
 DE Tryptic peptide from bovine serum phospholipase-D.
 KW Tryptic peptide; glycosyl-phosphatidylinositol; phospholipase-D;
 KW bovine; serum; probe; hybridisation; liver; signal peptide;
 KW recombinant protein; anchor; vector; cloning; protein secretion;
 KW fusion protein; purification.
 SO Bos taurus.

PN US5418147-A.
 PD 23-MAY-1995.
 PR 27-SEP-1990; 588896.
 PR 27-SEP-1990; US-588896.
 PR 31-MAR-1992; US-860825.
 PA (HOFF) HOFFMANN LA ROCHE INC.
 PI Huang K, Kochan JP, Li SH, Pan YE, Scallon BJ;
 PI Tsang TCH;
 DR WPI: 95-199737/26.
 PT New glycosyl-phosphatidyl:inositol-specific phospholipase D gene -
 PT used partic. for producing secreted recombinant proteins from
 PT GPI-anchored proteins expressed by cells
 PS Example 1; Column 19-20; 54pp; English.
 CC The sequence represents a fraction-T34 tryptic peptide derived from
 CC bovine serum glycosyl-phosphatidylinositol (GPI)-specific
 CC phospholipase-D. The sequence, along with those of peptides
 CC R75114-26, is used in construction of probes for isolation of the
 CC corresponding full-length bovine liver gene (encoding R75109). The
 CC resulting gene may be connected to a heterologous gene linked at the
 CC 3'-end to a C-terminal signal peptide (e.g. R75129 or R75130) gene to
 CC form a GPI-anchored molecule, where the C-terminal peptide directs
 CC attachment of a GPI anchor to the heterologous protein. The resulting
 CC construct may be inserted in a vector for expression in a host cell.
 CC The enzyme selectively hydrolyses the inositol-phosphate linkage of
 CC GPI-anchored proteins, GPI lipids and related molecules, resulting in
 CC direct secretion into the medium and simple recombinant protein
 CC purification.
 SO Sequence 15 AA;

Query Match 69.8%; Score 37; DB 16; Length 15;
 Best Local Similarity 71.4%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 7 tlnhngs 13
 QY 1 TLHONGNV 7

RESULT 11
 ID R47237 standard; Protein: 532 AA.
 AC R47237:
 DT 06-SEP-1994 (first entry)
 DE Wild-type Feline Herpes Virus ORF-2-encoded protein.
 KW Feline herpes virus; FHV; genome; mutant; vaccine; ORF-2;
 KW feline viral rhinotracheitis; herpesviridae family;
 OS Feline Herpes Virus.
 PN WO9403621-A.
 PD 17-FEB-1994.
 PF 23-JUL-1993; E01971.
 PR 30-JUL-1992; EP-202365.
 PA (ALKO) AKZO NV.
 PI Sondermeijer PJA, Willemsse MJ;
 PI WPI: 94-065709/08.
 DR N-PSDB: 056188.
 PT Feline herpes virus mutant comprising a heterologous gene
 PT inserted in the virus genome - for vaccination against feline
 PT herpes virus and other feline pathogens.
 PS Claim 2; Page 39-40; 55pp; English.
 CC Mutant versions of the wild-type feline herpes virus genome (056188)
 CC are claimed. Esp. the FHV mutant is one which fails to produce one
 CC or more antigenic or functional polypeptides. The mutant may contain
 CC a heterologous nucleic acid sequence inserted within one of the 6
 CC open reading frames. The attenuated FHV mutants can be used to prepare
 CC vaccines against feline viral rhinotracheitis and, where the
 CC heterologous insert encodes an antigen of a feline pathogen,
 CC against other infectious diseases of feline.
 SO Sequence 532 AA;

Query Match 69.8%; Score 37; DB 9; Length 532;
 Best Local Similarity 57.1%; Pred. No. 2.96e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 147 tlhngd 153
 |||:|:
 QY 1 TLHONGN 7

RESULT 12
 ID R96588 standard; Protein: 532 AA.
 AC R96588;

DE 17-DEC-1996 (first entry)
 DE feline herpesvirus glycoprotein E.
 DE feline herpesvirus; FHV; deletion; insertion; glycoprotein E;
 KW gp E; unique short region; attenuated virus; vaccine; protection;
 KW cat; vector; delivery; antigen.
 OS feline herpesvirus.

PN MO9613575-A1.
 PN 09-MAY-1996.
 PF 26-OCT-1995; U13975.
 PR 26-OCT-1994; US-329883.
 PA (SYTR) SYNTRO CORP.
 PI Cochran MD. McDonnell MW;
 DR WPI: 96-233489/24.
 DR N-PSDB; T28190.

PT Recombinant feline herpes virus attenuated by alteration of the gE
 PT gene - and related homology vectors, useful in vaccines, also as
 PT vectors for delivering antigens or therapeutic agents to mammals and
 PT birds

PS Example 1; Pages 76-78; 162pp; English.
 CC A novel feline herpesvirus (FHV) contains a deletion or insertion
 CC in the glycoprotein E (gp E) gene, in the unique short region of
 CC the viral genome. Altering the gp E gene, so that it no longer
 CC produces functional gp E.1.e. the present sequence, attenuates the
 CC virus. The attenuated virus is useful in vaccines to protect cats
 CC against FHV (and opt. against other diseases if appropriate DNA is
 CC inserted), and as a vector for delivering vaccinating antigens or
 CC therapeutic agents (e.g. antisense mols., ribozymes, interferon
 CC inducers, hormones, lymphokines, etc.) to other mammals (including
 CC humans) or birds.
 SQ Sequence 532 AA;

Query Match 69.8%; Score 37; DB 19; Length 532;
 Best Local Similarity 57.1%; Pred. No. 2.96e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 147 tlhngd 153
 |||:|:
 QY 1 TLHONGN 7

RESULT 13
 ID R71616 standard; Protein: 716 AA.

AC R71616;
 DT 10-NOV-1995 (first entry)
 DE Thermostable pululanase.
 DE pululanase; thermostable.

OS Thermus sp.
 PN J07023783-A.
 PN 27-JAN-1995.
 PD 15-JUL-1991; 200089.
 PF 15-JUL-1991; JP-200089.
 PA (NISO) NIPPON SHOKUHIN KAKO KK.
 PA (OPPP) OP CORP.
 PA (RIKA) RIKAGAKU KENKYUSHO.
 PA (SHKO) SHINGIUTSU JIGYODAN.
 DR WPI: 95-100945/14.
 DR N-PSDB; O85999.

PT A thermostable pululanase gene - for the production of
 PT recombinant pululanase free from amylase activity

PS Claim 1; Fig 2-4; 9pp; Japanese.
 CC The new thermostable pululanase gene (O85999) can be produced by
 CC recombinant DNA techniques. Esp. the pululanase prep. has no
 CC amylase activity.
 SQ Sequence 716 AA;

Query Match 69.8%; Score 37; DB 13; Length 716;
 Best Local Similarity 83.3%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 328 tlhng 333
 |||:|:
 QY 1 TLHONG 6

RESULT 14
 ID W37501 standard; Protein: 816 AA.
 AC W37501;

DE 20-APR-1998 (first entry)
 DE Human nel-related protein type 2.
 KW Human; foetal brain cDNA library; GDP dissociation stimulating protein;
 KW brain specific nucleosome assembly protein; diagnosis; therapy;
 KW skeletal muscle specific ubiquitin conjugating enzyme; RMP-2; NPK;
 KW nel-related protein type 1; nel-related type 2; hereditary disease;
 KW cancer.

OS Homo sapiens.
 PN EP-796913-A2.
 PD 24-SEP-1997.
 PF 19-MAR-1997; 104842.
 PR 05-MAR-1997; JP-069163.
 PR 19-MAR-1996; JP-063410.
 PA (SARA) OTSUKA PHARM CO LTD.
 PI Fujiwara T, Horie M, Watanabe T;
 DR WPI: 97-459830/43.
 DR N-PSDB; V01882, V01883.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -
 PT useful for diagnosis or therapy of hereditary disease and cancer
 PS Claim 22; Page 104-107; 123pp; English.
 CC The present sequence represents a nel-related protein type 2 isolated
 CC from a human foetal brain cDNA library. The nucleotide or amino acid
 CC sequences are useful for in-vitro diagnosis of hereditary diseases and
 CC cancer and for preparation of pharmaceuticals.
 SQ Sequence 816 AA;

Query Match 69.8%; Score 37; DB 27; Length 816;
 Best Local Similarity 83.3%; Pred. No. 2.96e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 701 lhng 706
 |||:|:
 QY 2 LHONGN 7

RESULT 15
 ID R75109 standard; Protein: 839 AA.

AC R75109;
 DT 15-MAY-1996 (first entry)
 DE Glycosyl-phosphatidylinositol-specific phospholipase-D.
 KW Glycosyl-phosphatidylinositol; phospholipase-D; bovine; liver;
 KW signal peptide; recombinant protein; metal ion binding domain;
 KW tryptic peptide; N-glycosylation; anchor; vector; cloning;
 KW fusion protein; protein secretion; purification.
 OS Bos taurus.

FH Key
 FT peptide 1.23 Location/Qualifiers
 FT peptide /note= "Signal peptide"
 FT peptide 24..40 /note= "N-terminal peptide homologous to R75112"
 FT peptide 36..50 /note= "Tryptic peptide R75119"
 FT peptide 56..70 /note= "Tryptic peptide R75120"
 FT peptide 56..77 /note= "Tryptic peptide R75121"
 FT peptide 126..138 /note= "Tryptic peptide R75116"
 FT peptide 236..261 /note= "Tryptic peptide R75113"
 FT peptide 349..358 /note= "Tryptic peptide R75113"

```

FT domain /note="Tryptic peptide R75114"
FT 379..402
FT /note="Putative metal ion binding domain"
FT 380..408
FT peptide /note="Tryptic peptide R75116"
FT 449..471
FT domain /note="Putative metal ion binding domain"
FT 449..477
FT peptide /note="Tryptic peptide Q75118"
FT 512..534
FT domain /note="Putative metal ion binding domain"
FT 623..636
FT peptide /note="Tryptic peptide R75115"
FT 678..692
FT domain /note="Tryptic peptide R75122"
FT 717..739
FT peptide /note="Putative metal ion binding domain"
FT 786..798
FT /note="Tryptic peptide R75123"
FT 799..807
FT peptide /note="Tryptic peptide R75124"

FT US5418147-A.
FT 23-MAY-1995.
FT 27-SEP-1990: 588896.
FT 27-SEP-1990: US-588896.
FT 31-MAR-1992: US-860825.
FT (HOF) HOFFMANN LA ROCHE INC.
FT Huang K, Kochan JP, Li SH, Pan YE, Scallion BJ;
FT Tsang TCH;
FT WPI: 95-199737/26.
FT N-PSDB: 090582.
FT New glycosyl-phosphatidylinositol-specific phospholipase D gene -
FT used partic. for producing secreted recombinant proteins from
FT GPI-anchored proteins expressed by cells
FT Claim 1: Fig 5; 54pp; English.
FT The sequence represents a bovine liver glycosyl-phosphatidylinositol
FT (GPI)-specific phospholipase-D. The sequence contains an
FT N-terminal signal peptide and 4 regions showing sequence similarity
FT to metal ion binding domains of Integrin-a subunits, with a core
FT sequence (R75127) matching the EF-hand consensus motif
FT characteristic of calcium and magnesium binding proteins. Tryptic
FT peptides R75113-26 have been derived from the native purified protein.
FT The sequence contains 8 putative N-glycosylation sites. A gene
FT encoding the enzyme may be connected to a heterologous gene linked at
FT the 3'-end to a C-terminal signal peptide (e.g. R75129 or R75130) gene
FT to form a GPI-anchored molecule, where the C-terminal peptide directs
FT attachment of a GPI anchor to the heterologous protein. The resulting
FT construct may be inserted in a vector for expression in a host cell.
FT The enzyme selectively hydrolyses the inositol-phosphate linkage of
FT GPI-anchored proteins, GPI lipids and related molecules, resulting in
FT direct secretion into the medium and simple recombinant protein
FT purification.
FT Sequence 839 AA:

```

```

Query Match 69.8%; Score 37; DB 16; Length 839;
Best Local Similarity 71.4%; Pred. No. 2.96e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
DB 684 tlnhgs 690
OY 1 TLHONGN 7

```

Search completed: Fri Apr 16 14:25:30 1999
Job time : 13 secs.

 WIREIMAGE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSEARCH protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:22:33 1999; MPSEARCH time 2.66 Seconds
 Tabular output not generated. 171.661 Million cell updates/sec

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 131
 Sequence: 1 FSEPMROYLVLSMOSQ 17

Scoring table:
 PAM 150
 Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 29.800; Variance 41.738; scale 0.714

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	100.0	176	1 PNOG_HUMAN	NOCICEPTIN PRECURSOR (3.70e-15
2	131	100.0	181	1 PNOG_RAT	NOCICEPTIN PRECURSOR (3.70e-15
3	131	100.0	187	1 PNOG_MOUSE	NOCICEPTIN PRECURSOR (3.70e-15
4	66	50.4	468	1 C15Y_GABEL	PROBABLE CITRATE SYNTH	6.54e-01
5	65	49.6	317	1 ACCO_PEA	1-AMINOCYCLOPROPANE-1-	9.90e-01
6	64	48.9	207	1 YE39_METUA	HYPOTHETICAL PROTEIN M	1.48e+00
7	62	47.3	320	1 ACC3_CUCME	1-AMINOCYCLOPROPANE-1-	3.35e+00
8	61	46.6	318	1 ACC1_CUCME	1-AMINOCYCLOPROPANE-1-	4.97e+00
9	61	46.6	733	1 MCMS_MOUSE	DNA REPLICATION LICENSE	4.97e+00
10	61	46.6	734	1 MCMS_HUMAN	DNA REPLICATION LICENSE	4.97e+00
11	61	46.6	894	1 YN86_YEAST	HYPOTHETICAL 99.5 KD P	4.97e+00
12	61	46.6	984	1 MY8A_DICDI	MYOSIN IA HEAVY CHAIN	4.97e+00
13	61	46.6	1189	1 PTHP_MOUSE	PROTEIN-TYROSINE PHOSP	4.97e+00
14	60	45.8	144	1 YPHP_BACSV	HYPOTHETICAL 15.9 KD P	7.36e+00
15	60	45.8	422	1 CAN2_RABIT	CALPAIN 2, LARGE (CATA	7.36e+00
16	60	45.8	700	1 CAN2_MOUSE	CALPAIN 2, LARGE (CATA	7.36e+00
17	60	45.8	700	1 CAN2_RAT	CALPAIN 2, LARGE (CATA	7.36e+00
18	60	45.8	700	1 CAN2_HUMAN	CALPAIN 2, LARGE (CATA	7.36e+00
19	60	45.8	1187	1 PTNE_HUMAN	PROTEIN-TYROSINE PHOSP	7.36e+00
20	60	45.8	4466	1 DYHC_ANTCR	DYNEIN BETA CHAIN, CIL	7.36e+00
21	60	45.8	4466	1 DYHC_ANTCR	DYNEIN BETA CHAIN, CIL	7.36e+00
22	59	45.0	89	1 RFAP_SALTY	LIPOPOLYSACCHARIDE COR	1.08e+01
23	59	45.0	321	1 ACCO_D1ACA	PROBABLE 1-AMINOCYCLOP	1.08e+01

24	59	45.0	367	1 TRMU_MYCGE	PROBABLE TRNA (5-METHY	1.08e+01
25	59	45.0	370	1 TRMU_MTCEN	PROBABLE TRNA (5-METHY	1.08e+01
26	59	45.0	433	1 C15Y_CHICK	CITRATE SYNTHASE, MITO	1.08e+01
27	59	45.0	433	1 ENOG_HUMAN	GAMMA ENOLASE (EC 4.2.	1.08e+01
28	59	45.0	464	1 C15Y_PIG	CITRATE SYNTHASE, MITO	1.08e+01
29	59	45.0	716	1 MCMS_XENLA	DNA REPLICATION LICENSE	1.08e+01
30	59	45.0	946	1 GLNE_ECOLI	GLUTAMATE-AMMONIA-LIGA	1.08e+01
31	58	44.3	315	1 ACC3_LYCES	1-AMINOCYCLOPROPANE-1-	1.59e+01
32	58	44.3	316	1 ACC2_LYCES	1-AMINOCYCLOPROPANE-1-	1.59e+01
33	58	44.3	319	1 ACC1_PETHY	1-AMINOCYCLOPROPANE-1-	1.59e+01
34	58	44.3	320	1 ACC3_PETHY	1-AMINOCYCLOPROPANE-1-	1.59e+01
35	58	44.3	371	1 IDH_METJA	ISOCITRATE DEHYDROGENA	1.59e+01
36	58	44.3	393	1 FTSM_STINT3	PROBABLE CELL DIVISION	1.59e+01
37	58	44.3	433	1 ENOG_RAT	GAMMA ENOLASE (EC 4.2.	1.59e+01
38	58	44.3	433	1 ENOG_MOUSE	GAMMA ENOLASE (EC 4.2.	1.59e+01
39	58	44.3	486	1 NAM9_YEAST	NAM9 PROTEIN PRECURSOR	1.59e+01
40	58	44.3	531	1 PEPE_ASPNG	SERINE-TYPE CARBOXYPEP	1.59e+01
41	58	44.3	714	1 CAN1_HUMAN	CALPAIN 1, LARGE (CATA	1.59e+01
42	58	44.3	1135	1 NINS_DROME	NINAC SHORT PROTEIN (E	1.59e+01
43	58	44.3	1246	1 NARG_ECOLI	RESPIRATORY NITRATE RE	1.59e+01
44	58	44.3	1501	1 NINL_DROME	NINAC LONG PROTEIN (EC	1.59e+01
45	57	43.5	700	1 CAN2_CHICK	CALPAIN 2, LARGE (CATA	2.31e+01

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	176 AA.
AC	013519;			
DT	01-NOV-1997 (REL. 35, LAST CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG).			
GN	PNOG OR OFO.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUTEROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
CC	EUTHERIA; PRIMATES.			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 96323281.			
RA	MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,			
RA	MEUNIER J.-C., PARMENTIER M.,			
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-BRAIN;			
RA	NOTACKER H.P., REINSCHIED R.R., MANSOUR A., HENNINGSEN R.A.,			
RA	MOMMA F.J. JR., WATSON S.J., CIVELLI O.,			
RL	SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBD DATA BANKS.			
RN	[3]			
RP	STRUCTURE BY NMR OF NOCICEPTIN.			
RX	MEDLINE: 97312464.			
RA	SALVADORI S., PICONE D., TANGREDI T., GUERRINI R., SPADACCINI R.,			
RA	LAZARUS L.H., REGOLI D., TEMUSSI P.A.,			
CC	BIOCHEN. BIOPHYS. RES. COMMON. 233:640-643(1997).			
CC	[1-] FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE			
CC	RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY			
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED			
CC	IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).			
CC	[1-] SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.			
CC	[1-] TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.			
CC	[1-] PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES			
CC	PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.			
CC	[1-] PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO			
CC	BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.			
CC	[1-] SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS			
CC	FAMILY.			
DR	EMBL: X97370; E244777; -			
DR	EMBL: X97367; E244785; -			
DR	EMBL: X97368; E244785; JOINED.			
DR	EMBL: U48263; G1185010; -			
DR	MMV: 601459; -			
DR	PROSITE: PS01252; OPIOIDS_PRECURSOR; 1.			

KW OPIOID PEPTIDE: NEUROTRANSMITTER; NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
 SQ SEQUENCE 176 AA; 20295 MW; 8055843A CRC32;
 Query Match 100.0%; Score 131; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 3,70e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 149 FSEFMROYLVLSMOSSQ 165
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 2
 ID PNOG RAT STANDARD; PRT; 181 AA.
 AC 062923: 064162;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (ORL1 RECEPTOR AGONIST).
 GN PNOG.
 OS RATTUS NORVEGICUS (RAT).
 OC EUTHERIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RC SEQUENCE FROM N.A.
 RA STRAIN-SPRAGUE-DAWLEY;
 RA NOTACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSEN R.A.,
 RA MONSMA F.J. JR., MATSON S.J., CIVELLI O.,
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBD DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [3]
 RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
 RX TISSUE-BRAIN;
 RA MEDLINE; 96013745.
 RA MEUNIER J.-C., MOLLEREAU C., TOLL L., SNAUDEAU C., MOISAND C.,
 RA ALVIERIE P., BUIROU J.-L., GUILLOT J.-C., FERRARA P.,
 RA MONSARAT B., MARAGUIL H., VASSART G., PARMENTIER M., COSTENTIN J.,
 RL NATURE 377:532-535(1995).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
 AND BRAIN, BEING MORE ABUNDANT IN THE HYPOTHALAMUS AND STRIATUM.
 CC -1- ALSO FOUND IN SMALL AMOUNTS IN OVARY.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PFM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 CC EMBL; U48262; G1185012; -.
 DR EMBL; X97375; E244795; -.
 DR EMBL; S79730; G1172239; -.
 DR PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.
 KM OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
 FT DOMAIN 109 120 2 X 6 AA TANDEM REPEATS OF D-A-E-P-X-A.

FT REPEAT 109 114 1.
 FT REPEAT 115 120 2.
 SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;
 Query Match 100.0%; Score 131; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 3,70e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 154 FSEFMROYLVLSMOSSQ 170
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 3
 ID PNOG MOUSE STANDARD; PRT; 187 AA.
 AC 064387: 061105; 061938;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (N23K / N27K).
 GN PNOG OR NPNC1.
 OS MUS MUSCULUS (MOUSE).
 OC EUTHERIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 RN EUTHERIA; RODENTIA.
 RC SEQUENCE FROM N.A.
 RA TISSUE-BRAIN;
 RX MEDLINE; 96216718.
 RA HOULTANT T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.,
 RA BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96106851.
 RA SAITO Y., MARUYAMA K., SAIDO T.C., KAWASHIMA S.,
 RA BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-129/SV;
 RC MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE; 96279082.
 RA SAITO Y., MARUYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KAWAMURA K., SAIDO T.C., KAWASHIMA S.,
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN [5]
 RP SEQUENCE OF 79-187 FROM N.A.
 RX TISSUE-BRAIN;
 RX MEDLINE; 96207555.
 RA PAN Y.-X., XU J., PASTERNAK G.W.,
 RL BIOCHEM. J. 315:11-13(1996).
 CC -1- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC -1- DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC -1- LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC -1- PFM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS

CC FAMILY.
 DR EMBL: D82866; D1012281; -
 DR EMBL: D50056; D1009395; -
 DR EMBL: X57373; E244874; -
 DR EMBL: X57371; E244782; -
 DR EMBL: X57372; E244782; JOINED.
 DR EMBL: U44027; G1335870; -
 DR EMBL: D50055; G1304162; -
 DR MGD: MG1:105308; NPNC1.
 DR PROSITE: PS01252; OPIOIDS_PRECURSOR; 1.
 DR OPIOD PEPTIDE; NEUROTANSMITTER; NEUROPEPTIDE; ALTERNATIVE SPLICING;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 98 138
 FT PEPTIDE 141 157
 FT PEPTIDE 160 176
 FT DOMAIN 109 126
 FT REPEAT 109 114
 FT REPEAT 115 120
 FT REPEAT 121 126
 FT REPEAT 121 126
 FT VARSPLIC 186 187
 FT NV -> IOYIPRTACVSKTCRPGVRIIPSPRH (1N
 FT LONG FORM; N27K).
 FT V -> T (IN REF. 5).
 SQ CONFLICT 80 80
 SQ SEQUENCE 187 AA; 20884 MW; F810ABF0 CRC32;
 Query Match 100.0%; Score 131; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 3,70e-15;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 160 FSEFMROYLVLSMOSSQ 176
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 4
 ID C15Y CAEEL STANDARD; PRT; 468 AA.
 AC P34575;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE PROBABLE CITRATE SYNTHASE, MITOCHONDRIAL-precursor (EC 4.1.3.7).
 GN T20G5.2.
 OS CAENORHABDITIS ELIGANS.
 CC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA BERKS M., SMITH A.;
 RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: CITRATE + COA -> ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX (BY SIMILARITY).
 CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF OXIDATIVE
 CC METABOLISM.
 CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
 DR EMBL: Z20423; G458482; -
 DR PIR: S42370; S42370.
 DR HSSP: P23007; 1CSC.
 DR WORMPEP: T20G5.2; CE00513.
 DR PROSITE: PS00460; CITRATE_SYNTHASE; 1.
 KM HYPOTHETICAL PROTEIN; LYASE; TRICARBOXYLIC ACID CYCLE; MITOCHONDRION;
 KW TRANSIT PEPTIDE.
 FT CHAIN 1 468
 FT ACT_SITE 303 303
 FT ACT_SITE 349 349
 FT ACT_SITE 404 404
 FT ACT_SITE 404 404
 FT SEQUENCE 468 AA; 51540 MW; ADA3A630 CRC32;
 Query Match 50.4%; Score 66; DB 1; Length 468;
 Best Local Similarity 63.6%; Pred. No. 6.54e-01;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 253 FAEIMRLIYVI 263
 QY 1 FSEFMROYLVLSMOSSQ 11
 RESULT 5
 ID ACCO_PEA STANDARD; PRT; 317 AA.
 AC P31239;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE (ACC OXIDASE) (ETHYLENE-
 DE FORMING ENZYME) (EFE).
 OS PISUM SATIVUM (GARDEN PEA).
 CC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; FABALES;
 CC FABACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, ALASKA; TISSUE-SHOOT;
 RX MEDLINE: 94105307.
 RA PECK S.C., OLSON D.C., KENDE H.;
 RL PLANT PHYSIOL. 101:689-690(1993).
 CC -1- CATALYTIC ACTIVITY: 1-AMINOCYCLOPROPANE-1-CARBOXYLATE + O2 -
 CC ETHYLENE + HCN + CO(2) + 2 H(2)O.
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ETHYLENE.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING FRUIT RIPENING.
 CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF
 CC OXIDOREDUCTASES.
 DR EMBL: M98357; G169041; -
 DR FRUIT RIPENING; ETHYLENE BIOSYNTHESIS; OXIDOREDUCTASE; IRON;
 KW VITAMIN C; MULTIGENE FAMILY.
 SQ SEQUENCE 317 AA; 36053 MW; 627C7319 CRC32;
 Query Match 49.6%; Score 65; DB 1; Length 317;
 Best Local Similarity 29.4%; Pred. No. 9.50e-01;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 280 FDDMYKMGKRFQAKE 296
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 6
 ID YE39_METJA STANDARD; PRT; 207 AA.
 AC O58834;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN MJ1439.
 GN MJ1439.
 OS METHANOCOCCUS JANNASCHII.
 CC ARCHAEABACTERIA; EUVRYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLADER A.,
 RA SCOTT J.L., GEORGHAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., MOSE C.R., VENTER J.C.;
 RA SCIENCE 273:1058-1073(1996).
 CC -1- SIMILARITY: BELONGS TO THE THERMONUCLEASE FAMILY.
 DR EMBL: U67584; G1592087; -
 DR TIGR: MJ1439; -
 DR PROSITE: PS01123; TNASE_1; 1.
 DR PROSITE: PS01284; TNASE_2; 1.

KM HYPOTHETICAL PROTEIN; HYDROLASE; NUCLEASE; ENDONUCLEASE.
 FT ACT_SITE 70 70 BY SIMILARITY.
 FT ACT_SITE 78 78 BY SIMILARITY.
 FT ACT_SITE 148 148 BY SIMILARITY.
 SQ SEQUENCE 207 AA; 24547 MW; E0CF3191 CRC32;
 Query Match 48.9%; Score 64; DB 1; Length 207;
 Best Local Similarity 60.0%; Pred. No. 1.49e+00;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 6 IMRKLFLISM 15
 QY 4 FMRQYLVLSM 13
 RESULT 7
 ID ACC3-CUCME STANDARD; PRT; 320 AA.
 AC P54847;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 3 (ACC OXIDASE) (ETHYLENE-FORMING ENZYME) (EFE).
 DE FORMING ENZYME (EFE).
 GN AC03.
 OS CUCUMIS MELO (MUSKMELO).
 OC EUPAROTIA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 CC VIOLAE; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-LEAF;
 RX LASSERE E., BOUQUIN T., HERNANDEZ J.A., BULL J., PECH J.C.,
 RA BALAGUE C.;
 RL MOL. GEN. GENET. 251:81-90(1996).
 CC -1- CATALYTIC ACTIVITY: 1-AMINOCYCLOPROPANE-1-CARBOXYLATE + O2 = ETHYLENE + HCN + CO(2) + 2 H(2)O.
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ETHYLENE.
 CC -1- TISSUE SPECIFICITY: FLOWERS.
 CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF OXIDOREDUCTASES.
 DR EMBL: X95553; E221407; -
 KW FRUIT RIPENING; ETHYLENE BIOSYNTHESIS; OXIDOREDUCTASE; IRON;
 KM VITAMIN C; MULTIGENE FAMILY.
 SQ SEQUENCE 320 AA; 36397 MW; F3FD4947 CRC32;
 Query Match 47.3%; Score 62; DB 1; Length 320;
 Best Local Similarity 29.4%; Pred. No. 3.35e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 282 FDDYMKLVGLKFOAKE 298
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 8
 ID ACC1-CUCME STANDARD; PRT; 318 AA.
 AC Q04644;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE 1 (ACC OXIDASE) (ETHYLENE-FORMING ENZYME) (EFE) (PMEI).
 GN AC01.
 OS CUCUMIS MELO (MUSKMELO).
 OC EUPAROTIA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 CC VIOLAE; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-FRUIT;
 RX MEDLINE; 93185633.
 RA BALAGUE C., WATSON C.F., TURNER A.J., ROUGE P., PICTON S.,
 PECH J.C., GRIERSON D.;

RL EUR. J. BIOCHEM. 212:27-34(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. ANDRES; TISSUE-FRUIT;
 RA YAMAMOTO M., MIKI T., ISHII Y., NAKAGAWA H., OGURA N., SATO T.;
 RN SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CANTALOUPE CHARENTAIS; TISSUE-LEAF;
 RX MEDLINE; 96213994.
 RA LASSERE E., BOUQUIN T., HERNANDEZ J.A., BULL J., PECH J.C.,
 RA BALAGUE C.;
 RL MOL. GEN. GENET. 251:81-90(1996).
 CC -1- CATALYTIC ACTIVITY: 1-AMINOCYCLOPROPANE-1-CARBOXYLATE + O2 = ETHYLENE + HCN + CO(2) + 2 H(2)O.
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ETHYLENE.
 CC -1- TISSUE SPECIFICITY: FRUIT.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING FRUIT RIPENING.
 CC -1- SIMILARITY: BELONGS TO THE IRON/ASCORBATE-DEPENDENT FAMILY OF OXIDOREDUCTASES.
 DR EMBL: X69955; G22663; -
 DR EMBL: D31727; G695400; -
 DR EMBL: X95551; E221404; -
 DR PIR: S29395; S29395.
 KW FRUIT RIPENING; ETHYLENE BIOSYNTHESIS; OXIDOREDUCTASE; IRON;
 KM VITAMIN C; MULTIGENE FAMILY.
 SQ SEQUENCE 318 AA; 36127 MW; 469A2496 CRC32;
 Query Match 46.6%; Score 61; DB 1; Length 318;
 Best Local Similarity 29.4%; Pred. No. 4.97e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 284 FEDYMKLVGLKFOAKE 300
 QY 1 FSEFMROYLVLSMOSSQ 17
 RESULT 9
 ID MCM5 MOUSE STANDARD; PRT; 733 AA.
 AC P49718;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE DNA REPLICATION LICENSING FACTOR MCM5 (CDC46 HOMOLOG) (P1-CDC46).
 GN MCM5 OR MCM5 OR CDC46.
 OS MUS MUSCULUS (MOUSE).
 OC EUPAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 95334361.
 RA KIMURA H., TAKIZAWA N., NOZAKI N., SUGIMOTO K.;
 RL NUCLEIC ACIDS RES. 23:2097-2104(1995).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: BELONGS TO THE MCM FAMILY.
 DR EMBL: D26090; G940404; -
 DR MGD; MGI:103197; MCM5.
 DR PROSITE; PS00847; MCM1; 1.
 DR PROSITE; PSS0051; MCM2; 1.
 KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;
 KW DNA REPLICATION; ATP-BINDING.
 FT DOMAIN 330 536 MCM.
 FT NP_BIND 380 387 ATP (POTENTIAL).
 SQ SEQUENCE 733 AA; 82342 MW; DA793E5F CRC32;
 Query Match 46.6%; Score 61; DB 1; Length 733;
 Best Local Similarity 70.0%; Pred. No. 4.97e+00;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 35 FKEFLROYRY 44
 QY 1 FSEFMROYLV 10

RESULT 10
ID MCM5_HUMAN STANDARD; PRT; 734 AA.
AC P33992; Q14578;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE DNA REPLICATION LICENSING FACTOR MCM5 (CDC46 HOMOLOG) (P1-CDC46).
GN MCM5 OR CDC46.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA HU B.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RM [2]
RP SEQUENCE FROM N.A.
RA MINURA S., NISHIMOTO S., KUBOTA Y., TAKISAWA H., NOJIMA H.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 352-590 FROM N.A.
RC TISSUE-CERVIX;
RX MEDLINE: 94089373.
RA HU B., BURKHART R., SCHULTE D., MUSHL C., KNIPPERS R.;
RL NUCLEIC ACIDS RES. 21:5289-5293(1993).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE MCM FAMILY.
DR EMBL: X74795; G895843; -;
DR EMBL: D83986; G1237079; -;
DR PIR: S43199; S43199.
DR PROSITE: PS00847; MCM_1; 1.
DR PROSITE: PS50051; MCM_2; 1.
KM TRANSCRIPTION REGULATION: DNA-BINDING; NUCLEAR PROTEIN;
KM DNA REPLICATION: ATP-BINDING.
FT DOMAIN 331 537 MCM.
FT NP_BIND 381 388 ATP (POTENTIAL).
FT CONFLICT 180 180 S -> T (IN REF. 2).
FT CONFLICT 241 241 V -> M (IN REF. 2).
FT CONFLICT 434 434 M -> M (IN REF. 2).
FT CONFLICT 527 527 E -> V (IN REF. 2).
FT CONFLICT 593 603 AROHERSDRR -> PSTSTGYTA (IN REF. 2).
FT CONFLICT 612 613 QL -> HV (IN REF. 1).
FT CONFLICT 717 717 R -> P (IN REF. 1).
SQ SEQUENCE 734 AA; 82308 MM; 3D6479B2 CRC32;
Query Match 46.6%; Score 61; DB 1; Length 734;
Best Local Similarity 70.0%; Pred. No. 4.97e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 35 EKEFLQRYV 44
Y 1 1 1 1 1
OY 1 FSEFMROYLV 10

RESULT 11
ID YN86_YEAST STANDARD; PRT; 894 AA.
AC P27514;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOHETICAL 99.5 KD PROTEIN IN URK1-SM1 INTERGENIC REGION.
GN YNR013C OR N2052.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE: 95208356.
RA VERHAASELT P., AERT R., VOET M., VOLCKAERT G.;
RL YEAST 10:1355-1361(1994).
RN [2]

RP SEQUENCE OF 1-293 FROM N.A.
RA MAURER C.T.C., URBANUS J.H.M., PLANTA R.J.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RM [3]
RP SEQUENCE OF 784-894 FROM N.A.
RC STRAIN-ATCC 28383 / FL100;
RX MEDLINE: 90384830.
RA KERN L.;
RL NUCLEIC ACIDS RES. 18:5279-5279(1990).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NAD(P)H/PHOS7 FAMILY OF TRANSPORTERS.
CC PHOS7 SUBFAMILY.
DR EMBL: X77395; G496729; -;
DR EMBL: Z71628; E239581; -;
DR EMBL: Z71629; E328814; -;
DR EMBL: X53998; E32709; ALT_INIT.
DR PIR: S29375; S29375.
DR PIR: S45135; S45135.
KM HYPOTHETICAL PROTEIN: TRANSMEMBRANE.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 557 577 POTENTIAL.
FT TRANSMEM 602 622 POTENTIAL.
FT TRANSMEM 642 662 POTENTIAL.
FT TRANSMEM 682 702 POTENTIAL.
FT TRANSMEM 706 726 POTENTIAL.
FT TRANSMEM 738 758 POTENTIAL.
FT TRANSMEM 777 797 POTENTIAL.
FT TRANSMEM 799 819 POTENTIAL.
FT TRANSMEM 824 844 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
SQ SEQUENCE 894 AA; 99490 MM; 949FB214 CRC32;

Query Match 46.6%; Score 61; DB 1; Length 894;
Best Local Similarity 46.2%; Pred. No. 4.97e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

DB 735 FNNFMWTVLVAM 747
Y 1 1 1 1 1
OY 1 FSEFMROYLVLM 13

RESULT 12
ID MYSA_DICDI STANDARD; PRT; 994 AA.
AC P22467;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-NOV-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMIA.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; SARCODINA; RHIZOPODA;
OC EUMYCETAZOA; DICTYOSTELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A5-2;
RX MEDLINE: 92096486.
RA TITUS M.A., WARRICK H.M., SPUDICH J.A.;
RL CELL REGUL. 1:55-63(1989).
CC -1- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE, AND EXTENSION AND RETRACTION OF PSEUDOPODIA OR LAMELLIPODIA.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN TYPE I FAMILY (SMALL MYOSINS).
DR EMBL: S73909; G241268; -;
DR PIR: J00151; J00151.
DR DICTYDB: DDO1028; ABMA.
KM MYOSIN: ACTIN-BINDING; ATP-BINDING; MULTIGENE FAMILY.
FT DOMAIN 1 782 GLOBULAR HEAD-LIKE DOMAIN.
FT NP_BIND 105 112 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
ATP (POTENTIAL).

```

FT DOMAIN 575 655 ACTIN-BINDING.
FT DOMAIN 830 931 MEMBRANE-BINDING (POTENTIAL).
SQ SEQUENCE 994 AA; 113286 MW; 5609233F CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 1; Length 994;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 653 FSEFMROYLVLSM 665
1 FSEFMROYLVLSM 13

RESULT 13
ID PINE_MOUSE STANDARD; PRT; 1189 AA.
AC Q62130;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEIN-TYROSINE PHOSPHATASE PTP36 (EC 3.1.3.48).
GN PTPN14.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CB-17-SCID; TISSUE-THYMUS;
RX MEDLINE; 94354845.
RA SAWADA M., OGATA M., FUJINO Y., HAMAKA T.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 203:479-484(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF T CELL DEVELOPMENT.
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- TISSUE SPECIFICITY: THYMUS; IN CELLS OF BOTH HEMATOPOIETIC AND
CC NON-HEMATOPOIETIC ORIGINS.
CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EGRIN, MOESIN,
CC RADIXIN, AND TALIN.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
DR EMBL; D31844; G50731; -.
DR MGD; MGI:102467; PTPN14.
DR PROSITE; PS00660; BAND_4.1; 1.
DR PROSITE; PS00661; BAND_4.1.2; 1.
DR PROSITE; PS50057; BAND_4.1.3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR STRUCTURAL PROTEIN; CYTOSKELETON; HYDROLASE.
FT DOMAIN 75 239 BAND 4.1-LIKE DOMAIN.
FT DOMAIN 935 1189 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT SITE 1123 1123 BY SIMILARITY.
FT DOMAIN 566 573 POLY-PRO.
FT DOMAIN 635 639 POLY-GLY.
FT DOMAIN 712 718 POLY-GLU.
SQ SEQUENCE 1189 AA; 135030 MW; CED0FECA CRC32;

Query Match
Best Local Similarity 46.6%; Score 61; DB 1; Length 1189;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db 160 EFLREYLVFPM 171
1 EFLREYLVFPM 14

RESULT 14
ID YPHP_BACSU STANDARD; PRT; 144 AA.
AC P54170;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 15.9 KD PROTEIN IN ILVD-THYB INTERGENIC REGION.
GN YPHP.

```

```

OS BACILLUS SUBTILIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / MABURG;
RA SOROKIN A.V., AZEVEDO V., ZUNSTEIN E., GALLERON N., EHRLICH S.D.,
RA SERROR P.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- SIMILARITY: STRONG, TO B.SUBTILIS Y01M.
DR EMBL; U77246; G125636; -.
DR EMBL; Z59115; E1183633; -.
DR SUBTILIST; BG11614; YPHP.
GN HYPOTHETICAL PROTEIN.
SQ SEQUENCE 144 AA; 15875 MW; 18864F3F CRC32;

Query Match
Best Local Similarity 45.8%; Score 60; DB 1; Length 144;
Matches 7; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

Db 5 YEEYMQ-LVPMR 17
1 FSEFMROYLVLSM 14

RESULT 15
ID CAN2_RABIT STANDARD; PRT; 422 AA.
AC P06814;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CALPAIN 2, LARGE (CATALYTIC) SUBUNIT (EC 3.4.22.17) (CALCIUM-ACTIVATED
DE NEUTRAL PROTEINASE) (CAMP) (M-TYPE) (FRAGMENT).
GN CAPN2.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; LAGOMORPHA.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 86250902.
RA EMORI Y., KAWASAKI H., SUGIHARA H., IMAJOH S., KAWASHIMA S.,
RA SUZUKI K.;
RL J. BIOL. CHEM. 261:9465-9471(1986).
RN [2]
RX CALCIUM-BINDING DATA.
RX MEDLINE; 87279982.
RA MINAMI Y., EMORI Y., KAWASAKI H., SUZUKI K.;
RL J. BIOCHEM. 101:889-895(1987).
CC -1- FUNCTION: CALPAINS ARE CALCIUM-ACTIVATED NON-LYSOSOMAL THIOL-
CC PROTEASES.
CC -1- ENZYME REGULATION: CALPAIN II IS ACTIVATED BY MILLIMOLAR
CC CONCENTRATIONS OF CALCIUM.
CC -1- SUBUNIT: HETERODIMER OF A LARGE (CATALYTIC) AND A SMALL
CC (REGULATORY) SUBUNIT.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- IN MAMMALS THERE SEEM TO BE 3 DISTINCT FORMS: CALPAIN I (MICRO-
CC MOLES CA++ REQUIRING), CALPAIN II (MILI-MOLE CA++ REQUIRING),
CC AND CALPAIN P94. THE SMALL UNIT IS COMMON TO ALL FORMS.
CC -1- THIS PROTEIN SEEMS TO BIND TWO MOLES OF CALCIUM.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2; ALSO KNOWN AS THE
CC CALPAIN FAMILY OF THIOL PROTEASES.
DR EMBL; M13797; G165666; -.
DR PIR; B24815; B24815.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
GN HYPOTHETICAL.
FT NON_TER 1
FT DOMAIN 1 1
FT DOMAIN 49 49 II, THIOL PROTEASE.
FT DOMAIN 50 255 III.
FT DOMAIN 256 422 IV, CALCIUM-BINDING.
FT CA_BIND 307 318 SITE 1.

```

FT	CA BIND	337	348	SITE 2.
FT	DOMAIN	372	383	ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT	DOMAIN	402	413	ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
SO	SEQUENCE	422 AA:	49494 MM:	BIFDFFA CRC32;

Query Match 45.88; Score 60; DB 1; Length 422;
Best Local Similarity 62.58; Pred. No. 7.36e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db	50 FSDFLRHY	57
OY	1 FSEFMROY	8

Search completed: Fri Apr 16 14:22:41 1999
Job time : 8 secs.

THIS PAGE BLANK (USPTO)

 MWSEKELI
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:26:23 1999: MapPar time 2.36 Seconds
 Tabular output not generated. 90.843 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 20.317; Variance 18.409; scale 1.104

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	100.0	176	1	PNOC_HUMAN NOCICEPTIN PRECURSOR (1.74e-03
2	53	100.0	181	1	PNOC_HUMAN NOCICEPTIN PRECURSOR (1.74e-03
3	53	100.0	187	1	PNOC_MOUSE NOCICEPTIN PRECURSOR (1.74e-03
4	45	84.9	1051	1	UBA2_MOUSE UBIQUITIN-ACTIVATING E	3.94e-01
5	45	84.9	1051	1	UBA1_WHEAT UBIQUITIN-ACTIVATING E	3.94e-01
6	44	83.0	384	1	TEAP_ECOLI HYPOTHETICAL 43.6 KD P	7.46e-01
7	43	81.1	377	1	SENI_CANAL BETA-GLUCAN SYNTHESIS-	4.75e+00
8	41	77.4	261	1	YBA_ECOLI HYPOTHETICAL 28.7 KD P	4.75e+00
9	41	77.4	261	1	YBA_ECOLI HYPOTHETICAL 28.7 KD P	4.75e+00
10	41	77.4	508	1	FMS1_YEAST FMS1 PROTEIN.	4.75e+00
11	41	77.4	726	1	PROBABLE SERINE/THREON	4.75e+00
12	41	77.4	863	1	STP1_YEAST STP1 PROTEIN.	4.75e+00
13	40	75.5	298	1	YDH_ECOLI HYPOTHETICAL OXIDOREDU	8.61e+00
14	40	75.5	639	1	GYRB_HALSO DNA GYRASE SUBUNIT B (8.61e+00
15	39	73.6	147	1	SP22_BACME ANTI-SIGMA F FACTOR (S	1.54e+01
16	39	73.6	232	1	PRAY_ECOLI LIPOPOLYSACCHARIDE COR	1.54e+01
17	39	73.6	232	1	PRAY_ECOLI LIPOPOLYSACCHARIDE COR	1.54e+01
18	39	73.6	412	1	YADQ_ECOLI HYPOTHETICAL FIMBRIN-L	1.54e+01
19	39	73.6	462	1	ANFK_AZOVI NITROGENASE IRON-IRON	1.54e+01
20	39	73.6	720	1	KRE6_YEAST BETA-GLUCAN SYNTHESIS-	1.54e+01
21	39	73.6	723	1	MY5B_MOUSE MYOSIN 5B (MYOSIN-LIKE	1.54e+01
22	39	73.6	740	1	KRE6_CANAL BETA-GLUCAN SYNTHESIS-	1.54e+01
23	39	73.6	749	1	CATPA_ASPPU CATALASE A (EC 1.1.1.1.	1.54e+01

RESULT ID	1	PNOC_HUMAN	STANDARD:	PRT:	176 AA.
AC	013519;				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC).				
GN	PNOC OR OFQ.				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; PRIMATES.				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 96323281.				
RA	MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,				
RL	MEUNIER J.-C., PARENTIER M.,				
RL	PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).				
RN	(2)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-BRAIN;				
RA	NOTHACKER H.P., REINSCHIED R.K., MANSOUR A., HENNINGSSEN R.A.,				
RL	MONSMA F.J.J., JR., WATSON S.J., CIVELLI O.,				
RL	SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.				
RN	(3)				
RP	STRUCTURE BY NMR OF NOCICEPTIN.				
RX	MEDLINE: 97312464.				
RA	SALVADORI S., PIGONE D., TANCREDI T., GUERRINI R., SPADACCINI R.,				
RL	LAZARUS L.H., REGOLI D., TEMOSSI P.A.,				
RL	BIOCHEM. BIOPHYS. RES. COMMON. 233:640-643(1997).				
CC	- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE				
CC	RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY				
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED				
CC	IN NEURONAL DIFFERENTIATION AND DEVELOPMENT (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.				
CC	- TISSUE SPECIFICITY: PREDOMINANTLY IN THE BRAIN AND SPINAL CORD.				
CC	- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES AT PAIRED BASIC RESIDUES				
CC	PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.				
CC	- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO				
CC	BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.				
CC	- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS				
CC	FAMILY.				
DR	EMBL: X97370; E244777; -				
DR	EMBL: X97367; E244785; -				
DR	EMBL: X97368; E244785; JOINED.				
DR	EMBL: U48263; G1185010; -				
DR	MTM: 601459; -				
DR	PROSITE: PS01252; OPIOIDS_PRECURSOR: 1.				

OPPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KW CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 127 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 130 146 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 149 165 NEUROPEPTIDE 2 (PROBABLE).
 SQ SEQUENCE 176 AA; 20295 MW; 8055843A CRC32;
 Query Match 100.0%; Score 53; DB 1; Length 176;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 169 TLHONGNV 176
 QY 1 TLHONGNV 8
 RESULT 2
 ID PNOG RAT STANDARD; PRT; 181 AA.
 AC 062923; 064162;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (ORL1 RECEPTOR AGONIST).
 GN PNOG.
 OS RATUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RA NOTHACKER H. P., REINSCHIED R. K., MANSOUR A., HENNINGSEN R. A.,
 RA MONSMA F. J., JR., WATSON S. J., CIVELLI O.,
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [3]
 RP SEQUENCE OF 13-181 FROM N.A., AND SEQUENCE OF 135-151.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96013745.
 RA MEUNIER J.-C., MOLLEREAU C., TOLL L., SUAUDEAU C., MOISAND C.,
 RA ALVIERIE P., BUTOUR J.-L., GUILLEMET J.-C., FERRARA P.,
 RA MONSARRAT B., MAZARGUIL H., VASSART G., PARMENTIER M., COSTENTIN J.,
 RL NATURE 377:532-535(1995).
 RN [4]
 RP FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT.
 CC [1] SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC [1] TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE SPINAL CORD
 AND BRAIN, BEING MORE ABUNDANT IN THE HYPOTHALAMUS AND STRIATUM.
 CC [1] ALSO FOUND IN SMALL AMOUNTS IN OVARY.
 CC [1] PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC [1] PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC [1] SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 DR EMBL; 048262; G1185012; -
 DR EMBL; X97375; E244795; -
 DR EMBL; S79730; G1172239; -
 DR PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE;
 KW CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 98 132 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 135 151 NOCICEPTIN (ORPHANIN FQ).
 FT PEPTIDE 154 170 NEUROPEPTIDE 2 (PROBABLE).
 FT DOMAIN 109 120 2 X 6 AA TANDEM REPEATS OF D-A-E-D-X-A.

REPEAT 109 114 1.
 FT REPEAT 115 120 2.
 SQ SEQUENCE 181 AA; 20478 MW; 1640366F CRC32;
 Query Match 100.0%; Score 53; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 174 TLHONGNV 181
 QY 1 TLHONGNV 8
 RESULT 3
 ID PNOG MOUSE STANDARD; PRT; 187 AA.
 AC 064387; 061105; 061938;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOG) (N23K / N27K).
 GN PNOG OR NPNC1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96216718.
 RA HOTTANT T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.,
 RA BIOCHEM. BIOPHYS. RES. COMMUN. 219:714-719(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96106851.
 RA SAITO Y., MARYAMA K., SAIDO T.C., KAWASHIMA S.,
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 217:539-545(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RX MEDLINE; 96323281.
 RA MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,
 RA MEUNIER J.-C., PARMENTIER M.,
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RX MEDLINE; 96279082.
 RA SAITO Y., MARYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,
 RA KAWAMURA K., SAIDO T.C., KAWASHIMA S.,
 RL J. BIOL. CHEM. 271:15615-15622(1996).
 RN [5]
 RP SEQUENCE OF 79-187 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 96207555.
 RA PAN Y.-X., XU J., PASTERNAK G.W.,
 RL BIOCHEM. J. 315:11-13(1996).
 RN [6]
 RP FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE
 RECEPTOR (OPRL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY
 MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED
 IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED
 INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND
 DECREASES LOCOMOTOR ACTIVITY.
 CC [1] SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.
 CC [1] TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN
 KIDNEY AND SPLEEN.
 CC [1] DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14
 AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.
 CC [1] LEVELS DECREASE SIGNIFICANTLY IN ADULTS.
 CC [1] PTM: SPECIFIC ENZYMAIC CLEAVAGES AT PAIRED BASIC RESIDUES
 PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.
 CC [1] PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC [1] ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED
 BY ALTERNATIVE SPLICING.
 CC [1] SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS

CC FAMILY
 DR EMBL: D82866; D1012281; -
 DR EMBL: D50056; D1009395; -
 DR EMBL: X97373; E244874; -
 DR EMBL: X97371; E244782; -
 DR EMBL: X97372; E244782; JOINED.
 DR EMBL: U44027; G1335870; -
 DR EMBL: D50055; G1304162; -
 DR MGI: MGI:105308; NPNC1.
 DR PROSITE: PS0152; OPTOIDS_PRECURSOR; 1.
 DR OPTOIDS PEPTIDE; NEUROTRANSMITTER; NEUROPEPTIDE; ALTERNATIVE SPLICING;
 KM CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL; REPEAT.
 FT SIGNAL 1 19
 FT PEPTIDE 96 138 POTENTIAL
 FT PEPTIDE 141 157 NEUROPEPTIDE 1 (PROBABLE).
 FT PEPTIDE 160 176 NOCICEPTIN (ORPHANIN FQ).
 FT DOMAIN 109 126 NEUROPEPTIDE 2 (PROBABLE).
 FT REPEAT 109 114 3 X 6 AA TANDEM REPEATS OF D-A-E-P-G-A.
 FT REPEAT 115 120 1.
 FT REPEAT 121 126 2.
 FT VANSPLC 186 187 3.
 FT CONFLICT 80 80 NV -> IQVPTACVHSKTCRQGVRIPPSPRH (IN
 FT SEQUENCE 187 AA; 20884 MW; F810ABP0 CRC32;
 SQ
 Query Match 100.0%; Score 53; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 1.74e-03;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 180 TLHONGNV 187
 |||||
 QY 1 TLHONGNV 8

RESULT 4
 ID UBA2 WHEAT STANDARD; PRT: 1051 AA.
 AC P31251;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-ACTIVATING ENZYME E1 2.
 GN UBA2.
 OS TRITICUM AESTIVUM (WHEAT).
 OC EURARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC CYPERALES; GRAMINEAE.
 CC [1]
 RP SEQUENCE FROM N.A.
 RN MEDLINE: 92340519.
 RA HATFIELD P.M., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 267:14799-14803(1992).
 CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLYLATING WITH ATP ITS
 CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS
 RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING
 AN UBIQUITIN-E1 THIOLESTER AND FREE AMP.
 CC -1- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE, ALLOWING IT TO
 ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A NEW UBIQUITIN
 FORMING AN ADENYLYL INTERMEDIATE AS THE PREVIOUS ONE IS
 TRANSFERRED TO THE THIOL SITE.
 CC -1- THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.
 CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.
 CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.
 DR EMBL: M90663; G170684; -
 DR PROSITE: PS00836; UBIQUITIN_ACTIVAT_1; 1.
 DR PROSITE: PS00865; UBIQUITIN_ACTIVAT_2; 1.
 KW UBIQUITIN CONJUGATION; LIGASE; MULTIGENE FAMILY; REPEAT.
 FT ACT SITE 626 626 BY SIMILARITY.
 FT DOMAIN 56 605 2 APPROXIMATE REPEATS.
 FT REPEAT 56 194 1-1.
 FT REPEAT 453 605 1-2.
 SQ SEQUENCE 1051 AA; 116825 MW; F1376A64 CRC32;

Query Match 84.9%; Score 45; DB 1; Length 1051;
 Best Local Similarity 75.0%; Pred. No. 3.94e-01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 93 TLHDDGNV 100
 |||||
 QY 1 TLHONGNV 8

RESULT 5
 ID UBA1 WHEAT STANDARD; PRT: 1051 AA.
 AC P20973;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE UBIQUITIN-ACTIVATING ENZYME E1 1.
 GN UBA1.
 OS TRITICUM AESTIVUM (WHEAT).
 OC EURARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC CYPERALES; GRAMINEAE.
 CC [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RN STRAIN-CV, AUGUSTA;
 RX MEDLINE: 90368797.
 RA HATFIELD P.M., CALLIS J., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 265:15813-15817(1990).
 CC -1- MUTAGENESIS OF CYS, AND ACTIVE SITE.
 RN MEDLINE: 92340519.
 RA HATFIELD P.M., VIERSTRA R.D.;
 RL J. BIOL. CHEM. 267:14799-14803(1992).
 CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLYLATING WITH ATP ITS
 CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS
 RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING
 AN UBIQUITIN-E1 THIOLESTER AND FREE AMP.
 CC -1- PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.
 CC -1- SUBUNIT: MONOMER.
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC -1- THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE, ALLOWING IT TO
 ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A NEW UBIQUITIN
 FORMING AN ADENYLYL INTERMEDIATE AS THE PREVIOUS ONE IS
 TRANSFERRED TO THE THIOL SITE.
 CC -1- THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.
 CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.
 CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.
 DR EMBL: M55604; G170780; -
 DR PIR: A38373; A38373.
 DR PIR: A42873; A42873.
 DR PROSITE: PS00536; UBIQUITIN_ACTIVAT_1; 1.
 DR PROSITE: PS00865; UBIQUITIN_ACTIVAT_2; 1.
 KW UBIQUITIN CONJUGATION; LIGASE; MULTIGENE FAMILY; REPEAT.
 FT ACT SITE 626 626
 FT DOMAIN 56 605 2 APPROXIMATE REPEATS.
 FT REPEAT 56 194 1-1.
 FT REPEAT 453 605 1-2.
 SQ SEQUENCE 1051 AA; 117007 MW; 614F10DB CRC32;

Query Match 84.9%; Score 45; DB 1; Length 1051;
 Best Local Similarity 75.0%; Pred. No. 3.94e-01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 93 TLHDDGNV 100
 |||||
 QY 1 TLHONGNV 8

RESULT 6
 ID YEAP_ECOLI STANDARD; PRT: 384 AA.
 AC P76245; P94742; P97193;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 43.6 KD PROTEIN IN GAP-A-RND INTERGENIC REGION.

GN YEAP.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
NN ENTEROBACTERIACEAE.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97251358.
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,
RA ISONO K., KISAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,
RA MAKINO K., MIKI T., MIZOUCHI K., MORI H., MORI T., MOTOWURA K.,
RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RA SATO N., SANEPEI G., SEKI Y., SIVASUNDARAM S., TAKAMI H.,
RA TAKEDA J., TAKEMOTO K., WADA C., YAMAMOTO Y., HORIUCHI T.;
NN D-N RES. 3:379-392(1996).
[1] SIMILARITY: BELONGS TO THE YESE/YHDA/YHJK/YJCC FAMILY.
DR EMBL; AE000274; G1788095; -
DR EMBL; D90823; G1736420; ALT_INIT.
DR EMBL; D90824; G1736430; -
DR ECOGENE; EG13502; YEAP.
KM HYPOTHEICAL PROTEIN.
SQ SEQUENCE 384 AA: 43628 MW: 77AF634D CRC32;

Query Match	83.0%	Score 44: DB 1; Length 38;
Best Local Similarity	75.0%	Pred. No. 7.46e-01;
Matches	6; Conservative	2; Mismatches 0; Indels 0; Gaps 0;
Db	284 TLONGEVY 291	
OY	1 TLHONGNV 8	

ID	RESULT	7	STANDARD:	PRT:	737 AA.
AC	SKN1 CANAL				
AD	P87024:				
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	BETA-GLUCAN SYNTHESIS-ASSOCIATED PROTEIN SKN1.				
GN	SKN1.				
OS	CANDIDA ALBICANS (YEAST).				
OC	EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 97234650.				
RA	MIO T., YAMADA-OKASE T., YABE T., NAKAJIMA T., ARISAWA M.,				
RA	YAMADA-OKASE H.;				
RL	J. BACTERIOL. 179:2363-2372(1997).				
CC	FUNCTION: REQUIRED FOR SYNTHESIS OF THE MAJOR BETA-GLUCANS OF				
CC	THE CELL WALL (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.				
CC	- SIMILARITY: STRONG, TO KRE6.				
DR	EMBL: D88491: D1020371: -				
KW	GLYCOPROTEIN; TRANSMEMBRANE; CELL WALL; SIGNAL-ANCHOR.				
FT	DOMAIN	1	281		
FT	TRANSEM	282	302		
FT					
FT	DOMAIN	303	737		
FT	DOMAIN	56	60		
FT	DOMAIN	188	193		
FT	DOMAIN	194	198		
FT	CARBOHYD	403	403		
FT	CARBOHYD	449	449		
FT	CARBOHYD	490	490		
FT	CARBOHYD	565	565		
FT	CARBOHYD	590	590		
FT	CARBOHYD	718	718		
SEQUENCE	737 AA; 83708 MW; 2A896F1A CRC32;				

```

Query Match      81.1%      Score 43: DB 1; Length 737;
Best Local Similarity 75.0%; Pred. No. 1.4e+00;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

Db      635 ALHPNGNV 642
      :||| |||
Oy      1 TLHONGNV 8

```

RESULT	8	STANDARD;	PRT;	224	AA.
ID	SEF_BACSU				
AC	P39135;				
DT	01-FEB-1995	(REL. 31, CREATED)			
DT	01-FEB-1995	(REL. 31, LAST SEQUENCE UPDATE)			
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)			
DE	SEF PROTEIN.				
GN	SEF OR LPA-8.				
OS	BACILLUS SUBTILIS.				
OC	POKRAYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE				
LN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.				
RP	STRATN-168;				
RC	STRATN-168;				
RX	MEDLINE; 92212296.				
RA	NAKANO M.M., CORBELL N., BESSON J., ZUBER P.,				
RA	MOL. GEN. GENET. 232:313-321(1992).				
LN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-168 / JH624;				
RA	GRANDI G.;				
LN	[3]				
RP	SUBMITTED (APR-1992) TO EMBL/GENBANK/DBEJ DATA BANKS.				

Query Match	77.48;	Score 41;	DB 1;	Length 224;
Best Local Similarity	71.48;	Pred. No. 4.75e+00;		

EKGGLSPDLDSFSVRLHQDQVSIELPDSHSPCIKTYEVL
PGYMAVCAHPDEPDITMVSYEEL -> GRÖRLIASA
(IN NON SURFACTIN-PRODUCING STRAINS).

360	SEQUENCE	224 AA:	26135 MW:	47296063 Cnc232.
361	CONFLICT	118	119	IA -> MP (IN REF. 4).
362	CONFLICT	97	97	G -> C (IN REF. 2, 3 AND 5)
363	CONFLICT	22	22	T -> S (IN REF. 2, 3 AND 5)
364	CONFLICT	22	22	

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 172 LHONGOV 178

Qy 2 LHONGOV 8

RESULT 9 STANDARD: PRT: 261 AA.

AC P75713; P77128;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 28.7 KD PROTEIN IN GIP-PDRA INTERGENIC REGION.
GN YLBA OR GLXB6.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.

[1]
RC SEQUENCE FROM N.A.
RA STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

[2]
RP SEQUENCE FROM N.A.
RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KUROI O.,
RA LEW H., LIN D., NAMATH A., OENNER P., SCHRAMM S., DAVIS R.W.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

[3]
RP SEQUENCE FROM N.A.
RA STRAIN-ECL1;
RA OBRADORS N., CUSA E., BALDOMA L., BADIA J., AGUIJAR J.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AE000157; G1786725;
DR EMBL: U82664; G1773195;
DR ECOGENE: EG13622; YLBA.

KW HYPOTHETICAL PROTEIN.
FT CONFLICT 204 204 A -> T (IN REF. 2).
FT CONFLICT 225 261 GDYIFMGYSLQAGYGVGRGEAFSYISKDCNRDVEI ->
FT SEQUENCE 261 AA; 28730 MW; C156E345 CRC32;
SQ

Query Match 77.4%; Score 41; DB 1; Length 261;
Best Local Similarity 100.0%; Pred. No. 4.75e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 65 TLHONG 70
Qy 1 TLHONG 6

RESULT 10 STANDARD: PRT: 508 AA.

AC P50264;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE FMS1 PROTEIN.
GN FMS1 OR YMR020W OR YMR711.09.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

[1]
RP SEQUENCE FROM N.A.
RA POUSETT D., MARCIREAU M., KARST F.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

[2]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C / AB972;
RA LYE G., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MULTICOPY SUPPRESSOR OF PENICILLIN RESISTANCE.
CC -1- SIMILARITY: TO C.ALBICANS CBP1.

DR EMBL: X81848; E119324; -;
DR EMBL: Z49211; G798930; -;
DR SGD: L0003091; FMS1.
SQ SEQUENCE 508 AA; 57805 MW; 647D1742 CRC32;

Query Match 77.4%; Score 41; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 4.75e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 TLHONG 31
Qy 1 TLHONG 6

RESULT 11 STANDARD: PRT: 726 AA.

AC P25341;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PROBABLE SERINE/THREONINE-PROTEIN KINASE KIN82 (EC 2.7.1.-).
GN KIN82 OR YCR91W OR YCR1153.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92254506.
RA WILSON C., BERGANTINO E., LANFRANCHI G., VALLE G., CARIGNANI G.,
RA FRONTALI L.;
RL YEAST 8.71-77(1992).

CC -1- SIMILARITY: TO S.POMBE SPAC48.05.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES.
DR EMBL: X59720; E264713; -;
DR PIR: S22258; S22258.
DR HSSP: P05132; ICTP.
DR SGD: L0000906; KIN82.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 324 602 PROTEIN KINASE.
FT NP-BIND 330 338 ATP (BY SIMILARITY).
FT BINDING 353 353 ATP (BY SIMILARITY).
FT ACT_SITE 449 449 BY SIMILARITY.

SQ SEQUENCE 726 AA; 82122 MW; A80B175B CRC32;

Query Match 77.4%; Score 41; DB 1; Length 726;
Best Local Similarity 71.4%; Pred. No. 4.75e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 457 LHONGHV 463
Qy 2 LHONGHV 8

RESULT 12 STANDARD: PRT: 863 AA.

AC P32578;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE SIP1 PROTEIN.

GN SIP1 OR YDR422C OR D9461.11.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92358217.
RA YANG X., HUBBARD E.J.A., CARLSON M.;
RL SCIENCE 257:680-682(1992).
[2]
RP SEQUENCE FROM N.A.

RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
 RA BENO A., CARPENTER J., CHEN E., CHERRY J.N., CHONG E., DUNCAN M.,
 RA HOLWICK-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
 RA MOSELEY D., NARAHARA K., NAMATH A., OEFNER P., OH C., PETEL F.X.,
 RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOEGREN T., SHROFF N.,
 RA WINKANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
 RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: NOT KNOWN; INTERACTS WITH THE SNE1 PROTEIN KINASE.
 CC COULD BE AN ADAPTOR THAT PROMOTE THE ACTIVITY OF SNE1 TOWARDS
 CC SPECIFIC TARGETS.
 CC -1- PPM: PHOSPHORYLATED BY SNE1 IN VITRO.
 CC -1- SIMILARITY: BELONGS TO THE 5'-AMP-ACTIVATED PROTEIN KINASE, BETA
 CC SUBUNIT FAMILY.
 DR EMBL: M90531; G172604; -;
 DR EMBL: U33007; G927721; -;
 DR PIR: S41984; S41984.
 DR SGD: L0001890; S1P1.
 KM PHOSPHORYLATION.
 SQ SEQUENCE 863 AA; 96258 MW; 29D4E3DD CRC32;

Query Match 77.4%; Score 41; DB 1; Length 863;
 Best Local Similarity 62.5%; Pred. No. 4.75e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 754 TLHONGNV 761
 OY 1 TLHONGNV 8

RESULT 13
 ID YDHF_ECOLI STANDARD; PRT; 298 AA.
 AC P7187;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL OXIDOREDUCTASE IN SODC-NEMA INTERGENIC REGION.
 GN YDHF.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
 CC STRONG, TO B. SUBTILIS YCSN.
 DR EMBL: AE000260; G1787936; -;
 DR ECOGENE: EG13420; YDHF.
 KM HYPOTHETICAL PROTEIN: OXIDOREDUCTASE.
 FT ACT SITE 128
 SQ SEQUENCE 298 AA; 33626 MW; 22B2E26 CRC32;

Query Match 75.5%; Score 40; DB 1; Length 298;
 Best Local Similarity 71.4%; Pred. No. 8.61e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 146 LHOSGKV 152
 OY 2 LHONGNV 8

RESULT 14
 ID GYRB_HAUSQ STANDARD; PRT; 639 AA.
 AC P21558;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE DNA GYRASE SUBUNIT B (EC 5.99.1.3).
 GN GYRB.
 OS HALOERAX SP. (STRAIN AA 2.2).
 OC ARCHAEABACTERIA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE.
 RN (1)

RP SEQUENCE FROM N.A.
 RX MEDLINE: 91100352.
 RA HOLMES M.L., DYALLSMITH M.L.;
 RL J. BACTERIOL. 173:642-648(1991).
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 CC STRANDED DNA IN AN ATP-DEPENDANT MANNER AND ALSO CATALYZES THE
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
 CC BREAKAGE AND REJOINING, THE B CHAIN CATALYZES ATP HYDOLYSIS.
 CC -1- MOTATIONS IN DNA GYRASE RESULT IN NOVIOICIN RESISTANCE IN
 CC HALOPHILIC ARCHAEABACTERIA.
 DR EMBL: M38373; G149024; -;
 DR PIR: A39135; A39135.
 DR PROSITE: PS00177; TOPOISOMERASE II; 1.
 KM TOPOISOMERASE, ISOMERASE, ATP-BINDING.
 SQ SEQUENCE 639 AA; 71138 MW; FFA41166 CRC32;

Query Match 75.5%; Score 40; DB 1; Length 639;
 Best Local Similarity 50.0%; Pred. No. 8.61e+00;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 63 ALHEDGSV 70
 OY 1 TLHONGNV 8

RESULT 15
 ID SP22_BACME STANDARD; PRT; 147 AA.
 AC P35148;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ANTI-SIGMA F FACTOR (STAGE II SPOULATION PROTEIN AB).
 GN SPOI1AB.
 OS BACILLUS MEGATERIUM.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93003525.
 RA TAO Y.P., HUDSPETH D.S.S., VARY P.S.;
 RL BIOCHIMIE 74:695-704(1992).
 CC -1- FUNCTION: BINDS TO SIGMA F AND BLOCKS ITS ABILITY TO FORM AN RNA
 CC POLYMERASE HOLOENZYME (E-SIGMA F). PHOSPHORYLATES SPOI1AA ON A
 CC SERINE RESIDUE. THIS PHOSPHORYLATION MAY ENABLE SPOI1AA TO ACT AS
 CC AN ANTI-ANTI-SIGMA FACTOR THAT COUNTERACTS SPOI1AB AND THUS
 CC RELEASES SIGMA F FROM INHIBITION (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO RBSW.
 DR EMBL: X63757; G39653; -;
 DR PIR: S22178; S22178.
 DR PIR: B48402; B48402.
 KM SPOULATION; TRANSFERASE: KINASE.
 SQ SEQUENCE 147 AA; 16382 MW; CD51F1A5 CRC32;

Query Match 73.6%; Score 39; DB 1; Length 147;
 Best Local Similarity 62.5%; Pred. No. 1.54e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 69 TLHEDGV 76
 OY 1 TLHONGNV 8

Search completed: Fri Apr 16 14:26:31 1999
 Job time : 8 secs.

 NW5N5H
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 16 14:26:48 1999; Maspar time 4.55 Seconds
 87.470 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.dep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 165420 segs, 49795644 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

sptrembl6
 1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
 9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 19.822; Variance 18.814; scale 1.054

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	46	86.8	176	6	062647	NOCICEPTIN/ORPHANIN FQ
2	43	81.1	251	2	025087	CONSERVED HYPOTHETICAL
3	43	81.1	636	2	E77993	DNA TOPOISOMERASE II S
4	43	81.1	2150	5	044131	C08G9.2 PROTEIN.
5	42	79.2	771	5	022783	COSMID T25G12.
6	41	77.4	221	2	054410	GLXB6.
7	41	77.4	303	10	P92932	LECTIN RELATED PROTEIN
8	41	77.4	351	2	069578	PUTATIVE LONG CHAIN FA
9	41	77.4	507	13	057439	I012.
10	41	77.4	507	13	013020	AMINO ACID TRANSPORTER
11	41	77.4	1275	14	041977	TEGUMENT PROTEIN/FGARA
12	40	75.5	328	1	028652	CONSERVED HYPOTHETICAL
13	40	75.5	609	5	016938	F2G5.1 PROTEIN.
14	40	75.5	745	2	032144	YORC PROTEIN.
15	39	73.6	74	2	049718	B1549_F3_145.
16	39	73.6	158	2	046245	NITROGENASE (FRAGMENT)
17	39	73.6	192	2	007135	B1306_03 PROTEIN.
18	39	73.6	234	10	040590	TFHP-1 PROTEIN.
19	39	73.6	329	2	046084	NITROGENASE 3 SUBUNIT
20	39	73.6	439	1	029985	NADH OXIDASE (NOXA-1).

21	39	73.6	453	13	042115	ARX HOMEOPROTEIN.	3.15e+01
22	39	73.6	623	3	042626	PROTEIN KINASE NRC-2.	3.15e+01
23	39	73.6	720	3	006472	BETA-GLUCAN SYNTHESIS-	3.15e+01
24	39	73.6	765	5	021478	SIMILARITY OVER A SHOR	3.15e+01
25	39	73.6	880	11	P97693	P105 COACTIVATOR.	3.15e+01
26	39	73.6	885	4	013122	100 KDA COACTIVATOR.	3.15e+01
27	39	73.6	890	3	060040	100 KDA PROTEIN.	3.15e+01
28	39	73.6	1288	11	P97608	5-OXO-L-PROLINASE.	3.15e+01
29	38	71.7	80	14	055781	NON-STRUCTURAL PROTEIN	5.50e+01
30	38	71.7	238	1	027227	N5-METHYL-TERAHYDROME	5.50e+01
31	38	71.7	249	1	028475	CONSERVED HYPOTHETICAL	5.50e+01
32	38	71.7	325	2	067552	HYPOTHETICAL 36.2 KD P	5.50e+01
33	38	71.7	362	5	001493	SIMILARITY TO THE CATA	5.50e+01
34	38	71.7	471	11	035162	MICROSOMAL STRESS 70 P	5.50e+01
35	38	71.7	550	3	059725	HYPOTHETICAL 63.2 KD P	5.50e+01
36	38	71.7	585	13	057396	RSOX23.	5.50e+01
37	38	71.7	601	10	023522	TRIACYLGLYCEROL LIPASE	5.50e+01
38	38	71.7	636	11	035656	FILENIN.	5.50e+01
39	38	71.7	639	3	043113	ARTISLFRITASE.	5.50e+01
40	38	71.7	653	14	087039	SEVCPZ, COMPLETE GENOM	5.50e+01
41	38	71.7	669	11	054770	LENS FIBER CELL BEADED	5.50e+01
42	38	71.7	718	2	034587	AMTX PROTEIN.	5.50e+01
43	38	71.7	754	11	061123	MATERNAL EMBRYONIC MES	5.50e+01
44	38	71.7	1053	5	024310	POLYPROTEIN.	5.50e+01
45	38	71.7	1337	2	P76465	F1337.	5.50e+01

ALIGNMENTS

RESULT	ID	AC	PRELIMINARY:	PRT:	176 AA.
1	02647	062647			
DT	01-AUG-1998	(TREMBLREL. 07, CREATED)			
DT	01-AUG-1998	(TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)			
DE	NOCICEPTIN/ORPHANIN FQ PRECURSOR.				
OS	BOS TAURUS (BOVINE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; ARTIODACTYLA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	OKUDA-ASHITAKA E., MINAMI T., TACHIBANA S., YOSHIMURA Y.,				
RA	NISHITUCHI Y., KIMURA T., ITO S.;				
RL	NATURE 392:286-289(1998).				
DR	EMBL; AB005251; D1026347;				
SQ	SEQUENCE 176 AA; 20255 MW; D2C23826 CRC32;				
Query Match	86.8%;	Score 46;	DB 6;	Length 176;	
Best Local Similarity	100.0%;	Pred. No. 4.67e-01;			
Matches	7;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Db	169	TLHONGN 175			
QY	1	TLHONGN 7			
RESULT	2	PRELIMINARY:	PRT:	251 AA.	
ID	025087				
AC	025087				
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)			
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-JUN-1998	(TREMBLREL. 06, LAST ANNOTATION UPDATE)			
DE	CONSERVED HYPOTHETICAL PROTEIN.				
GN	HP0318.				
OS	HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).				
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;				
OC	AEROBIC; MOTILE, HELICAL AND/OR VIBRIOID.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	STRAIN-26695;				
RA	MEDLINE: 97394467.				
RA	TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,				
RA	FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,				

```

RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODER A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., MALLIN E.,
RA HAYES W.S., BORODOVSKIY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.,
RL NATURE 368:539-547(1997).
DR EMBL; AF000550; G2313418; -.
DR TIGR; HP0318; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 251 AA; 28507 MW; F7119AC2 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 2; Length 251;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 105 TLHPNGV 112
OY 1 TLHPNGV 8

RESULT 3
ID P77993; PRELIMINARY; PRT; 636 AA.
AC P77993;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
GN DNA TOPOISOMERASE II SUBUNIT B.
OS THERMOTOGA MARITIMA.
OC PROKARYOTA; NOT YET CLASSIFIED.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8;
RL MEDLINE; 97017137.
RA GUIPAUD O., LABEDAN B., FORTERRER P.;
RL GENE 174:121-128(1996).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
OF DOUBLE-STRANDED DNA.
DR EMBL; U49692; G1622792; -.
DR PROSITE; PS00177; TOPOISOMERASE-II; 1.
DR PFAM; PF00204; DNA_TOPOISOIL.
KW ISOMERASE; TOPOISOMERASE; ATP-BINDING.
SQ SEQUENCE 636 AA; 72459 MW; D9915C88 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 2; Length 636;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 60 TLHDSGV 67
OY 1 TLHONGV 8

RESULT 4
ID 044131; PRELIMINARY; PRT; 2150 AA.
AC 044131;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
GN C0869.2 PROTEIN.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RL MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,

```

```

RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA GEISEL C., STELLYES L.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF036687; G2662541; -.
DR PROSITE; PS00280; BPTI_KUNITZ; 1.
DR PROSITE; PS00317; 4-DISULFIDE_CORE; 9.
KW SERINE PROTEASE INHIBITOR.
SQ SEQUENCE 2150 AA; 234130 MW; 8231D798 CRC32;

Query Match
Best Local Similarity 81.1%; Score 43; DB 5; Length 2150;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1641 LHONGV 1647
OY 2 LHONGV 8

RESULT 5
ID 022783; PRELIMINARY; PRT; 771 AA.
AC 022783;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
GN T25G12.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC DU Z.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC WATERSTON R.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U43283; G1125834; -.
SQ SEQUENCE 771 AA; 83449 MW; 9B6C7ABA CRC32;

Query Match
Best Local Similarity 79.28%; Score 42; DB 5; Length 771;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 722 TLHONGV 729
OY 1 TLHONGV 1

```


OY 1 TLHONGNV 8

RESULT 6 PRELIMINARY: PRT: 221 AA.
 ID 054410:
 AC 054410:
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE GLXB6.
 GN GLXB6.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA OBRAODORS N., CUSA E., BALDOMA L., BADIA J., AGUILAR J.;
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U89279; G2735241; -;
 SO SEQUENCE 221 AA; 24306 MW; C91F86AA CRC32;

Query Match 77.4%; Score 41; DB 2; Length 221;
 Best Local Similarity 100.0%; Pred. No. 9.95e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 65 TLHONG 70
 |||||
 OY 1 TLHONG 6

RESULT 7 PRELIMINARY: PRT: 303 AA.
 ID P92932:
 AC P92932:
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE LECTIN RELATED PROTEIN PRECURSOR (FRAGMENT).
 OS ALTIUM SATIVUM (CARLIC).
 OC EUKARYOTA: PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 CC LILIALES; LILIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ROOTS;
 RX MEDLINE: 97188556.
 RA SNETS K., VAN DAMME E., VERHAERT P., BARRE A., VAN LEUVEN F.,
 RA PEDMANS W.J.;
 RL PLANT MOL. BIOL. 33:223-234(1997).
 DR EMBL: U58949; G1840051; -;
 KW SIGNAL; LECTIN.
 FT NON_TER 1
 FT SIGNAL <1 12 POTENTIAL,
 FT CHAIN 13 303 LECTIN RELATED PROTEIN.
 SQ SEQUENCE 303 AA; 32749 MW; 6D218BC0 CRC32;

Query Match 77.4%; Score 41; DB 10; Length 303;
 Best Local Similarity 71.4%; Pred. No. 9.95e+00;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 242 LQDQNV 248
 |||||
 OY 2 LQDQNV 8

RESULT 8 PRELIMINARY: PRT: 351 AA.
 ID 069578:
 AC 069578:
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE PUTATIVE LONG CHAIN FATTY ACID-COA LIGASE.
 GN MLCB268.30.
 OS MYCOBACTERIUM LEPRAE.
 OC PROKARYOTA: FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.

RN [1]
 RP SEQUENCE FROM N.A.
 RA BROWN D., CHURCHER C.M.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA PARKHILL J., BARRELL B.G., RAJANDREAN M.A.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
 RL MOL. MICROBIOL. 7:197-206(1993).
 DR EMBL: AL022602; E1287820; -;
 KW LIGASE.
 SO SEQUENCE 351 AA; 37624 MW; 76BC96A3 CRC32;

Query Match 77.4%; Score 41; DB 2; Length 351;
 Best Local Similarity 75.0%; Pred. No. 9.95e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 136 TLNDQNV 143
 |||||
 OY 1 TLHONGNV 8

RESULT 9 PRELIMINARY: PRT: 507 AA.
 ID 057439:
 AC 057439:
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE I012.
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
 OC EUKARYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SHI Y.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF01906; G2731774; -;
 SO SEQUENCE 507 AA; 55595 MW; 4581F63F CRC32;

Query Match 77.4%; Score 41; DB 13; Length 507;
 Best Local Similarity 100.0%; Pred. No. 9.95e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 28 LHQNGN 33
 |||||
 OY 2 LHQNGN 7

RESULT 10 PRELIMINARY: PRT: 507 AA.
 ID 013020:
 AC 013020:
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE AMINO ACID TRANSPORTER CHAIN ASUR4.
 GN ASUR4.
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
 OC EUKARYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA SPINDLER B., MASTROBERARDINO L., CUSTER M., VERREY F.;
 RL PELIGERS ARCH. 434:323-331(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA VERREY F.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;

RA VERREY F.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RA SHI X.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Y12716; E1289893;
 SQ SEQUENCE 507 AA; 55529 MW; CDF914A3 CRC32;
 Query Match 77.4%; Score 41; DB 13; Length 507;
 Best Local Similarity 100.0%; Pred. No. 9.95e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 28 LHONGN 33
 111111
 QY 2 LHONGN 7

RESULT 11
 ID 041977 PRELIMINARY; PRT; 1275 AA.
 AC 041977;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE TEGMENT PROTEIN/EGARAT.
 GN GAMAHY ORF75B.
 OS MURINE HERPESVIRUS 68.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; GAMAHHERPESVIRINAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WUMS;
 RX MEDLINE: 97366649.
 RA VIRGIN H.W. IV, LATREILLE P., WANSLEY P., HALLSWORTH K., WECK K.E.,
 RA DAL CANTO A.J., SPECK S.H.;
 RL J. VIROL. 71:5894-5904(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WUMS;
 RA LATREILLE P., WANSLEY P., WATKINSON R.H.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U97553; G2318001;
 SQ SEQUENCE 1275 AA; 141953 MW; 518C97E0 CRC32;

Query Match 77.4%; Score 41; DB 14; Length 1275;
 Best Local Similarity 62.5%; Pred. No. 9.95e+00;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1057 TLHERGV 1064
 111111
 QY 1 TLHONGNV 8

RESULT 12
 ID 028652 PRELIMINARY; PRT; 328 AA.
 AC 028652;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 GN CONSERVED HYPOTHETICAL PROTEIN.
 OS ARCHAEoglobus fulgidus.
 OC ARCHAEABACTERIA; EURYARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49556;
 RX MEDLINE: 98049343.
 RA KLEIN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KITCHOW K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYPIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.R., BADGER J.H., GLODER A., ZHOU L.,

RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RL NATURE 390:364-370(1997).
 DR EMBL: AE000990; G2648933;
 DR TIGR: AF1621;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 328 AA; 35412 MW; F803A6BD CRC32;

Query Match 75.5%; Score 40; DB 1; Length 328;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 198 TLHGTIV 205
 111111
 QY 1 TLHONGNV 8

RESULT 13
 ID 016938 PRELIMINARY; PRT; 609 AA.
 AC 016938;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE F26G5.1 PROTEIN.
 DE F26G5.1 PROTEIN.
 GN F26G5.1.
 OS CAENORHABDITIS ELEGANS.
 OC EUDARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
 RA BURTON J., CONNELL M., COPSET T., COOPER J., COULSON A., CRAYTON M.,
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
 RA RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SKALDON N., SMITH A.,
 RA SONNHAMMER E., STADEN R., SULLSTON J., THIERRY-MIEG J., THOMAS K.,
 RA VAUDIN N., VAUGHAN K., WATKINSON R., WATSON A., WEINSTOCK L.,
 RA WILKINSON-SPROAT J., WOHLDMANN P.;
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA SAMMONS L., WOHLDMANN P., BECK C.;
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF022974; G2384841;
 SQ SEQUENCE 609 AA; 68849 MW; 804D9451 CRC32;

Query Match 75.5%; Score 40; DB 5; Length 609;
 Best Local Similarity 62.5%; Pred. No. 1.78e+01;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 506 TLQDSNV 513
 111111
 QY 1 TLHONGNV 8

RESULT 14
 ID 032144 PRELIMINARY; PRT; 745 AA.
 AC 032144;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 GN YURC PROTEIN.
 OS BACILLUS SUBTILIS.
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AAEVERO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOUSIER L., BRANS A., BRAUN M., BRIGHELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
 RA ENITAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOLGER D., FRITZ C.,
 RA FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N., GHM S.Y.,
 RA GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G., GIUSEPPI G.,
 RA GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A., HILBERT H.,
 RA HOLSAPEL S., HOSONO S., HULLO M.F., ITRAYA M., JONES L., JORIS B.,
 RA KARAMATA D., KASAHARA Y., KLERR-BLANCARD M., KLEIN C., KOBAYASHI Y.,
 RA KOETTER P., KONIGSTEIN G., KROGH S., KUMANO M., KURITA K., LAPIDUS A.,
 RA LARDINOIS S., LAUBER J., LAZAREVIC V., LEE S.M., LEVINE A., LIU H.,
 RA MASUDA S., MAUEL C., MEDIGUE C., MEDINA N., MELLAO R.P., MIZUNO M.,
 RA MOESTL D., NAKAI S., NOBACK M., NOONE D., O'REILLY M., OGAWA K.,
 RA OGIMARA A., OUDEGA B., PARK S.H., PARO V., POHL T.M., PORTELELE D.,
 RA PORWOLIK S., PRESCOTT A.M., PRESCAN E., PUTIC P., PURNELLE B.,
 RA RAPOPORT G., REY M., REYNOLDS S., RIEGER M., RIVOLTA C., ROCHA E.,
 RA ROCHE B., ROSE M., SADAIE Y., SATO T., SCAMIAN E., SCHLEICH S.,
 RA SCHROETER R., SCOFFONE F., SEKIGUCHI J., SEKOWSKA A., SEROR S.J.,
 RA SEROR P., SHIN B.S., SOLDO B., SOROKIN A., TACCONI E., TAKAGI T.,
 RA TAKAHASHI H., TAKEWAKU K., TAKEUCHI M., TAKAKOSHI A., TANAKA T.,
 RA TERPSTRA P., TOGNONI A., TOSATO V., UCHIYAMA S., VANDENBOL M.,
 RA VANNIER F., VASSAROTTI A., VIARI A., WABUTT R., WEDLER E., WEDLER H.,
 RA WEITZENEGGER T., WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K.,
 RA YASUMOTO K., YATA K., YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E.,
 RA YOSHIKAWA H., DANCHIN A.;
 RL NATURE 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: 299120; E1184327;
 SQ SEQUENCE 745 AA; 80433 MW; D96C4EAB CRC32;

Query Match 75.5%; Score 40; DB 2; Length 745;
 Best Local Similarity 57.1%; Pred. No. 1.78e+01;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 140 LHEDGNT 146
 QY 2 LHONGNV 8

RESULT 15
 ID Q49718 PRELIMINARY; PRT; 74 AA.
 AC Q49718;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
 DE B1549_F3_145;
 OS MYCOBACTERIUM LEPRAE;
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ROBISON K.;
 RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA ROBISON K.;
 RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA SMITH D.R.;
 RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA ROBISON K.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U00014; G466929;
 SQ SEQUENCE 74 AA; 7972 MW; 56A2486D CRC32;
 Query Match 73.6%; Score 39; DB 2; Length 74;
 Best Local Similarity 71.4%; Pred. No. 3.15e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 48 TLHEHGN 54
 QY 1 TLHONGN 7

Search completed: Fri Apr 16 14:27:20 1999
 Job time : 32 secs.

THIS PAGE BLANK (USPTO)

 [W] [O] [R] [L] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:33:06 1999: MasPar time 99.54 Seconds

Tabular output not generated. 431.758 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 100
 N.A. Sequence: 1 ACNTYNCAYCARAAAYGNAAYGTN 24
 Comp: TGNRANGTRGTYTRCMTTCRAN

Scoring table: TABLE backtranslated
 Gap 40

Mmatch STD: Dbase 0; Query 0

Searched: 2275026 segs, 89538244 bases x 2

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: emb1-ests16
 1:em-est1.2:em-gss1 3:em-gss2 4:em-gss3
 genbank-est109
 5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13
 10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17
 14:gb-est18 15:gb-est19 16:gb-est2 17:gb-est20
 18:gb-est21 19:gb-est3 20:gb-est4 21:gb-est5 22:gb-est6
 23:gb-est7 24:gb-est8 25:gb-est9 26:gb-gss1 27:gb-gss2
 28:gb-gss3 29:gb-gss4

Statistics: Mean 40.359; Variance 39.952; scale 1.010

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C	1	95	95.0	402	15	A1007467	L0-370M13R Ice plant L	3.12e-04
C	2	95	95.0	431	14	AA962868	L30-398T3 Ice plant La	3.12e-04
C	3	95	95.0	471	14	AA962918	L30-449T3 Ice plant La	3.12e-04
C	4	95	95.0	472	14	AA887369	L30-342T3 Ice plant La	3.12e-04
C	5	95	95.0	482	14	AA962885	L30-415T3 Ice plant La	3.12e-04
C	6	95	95.0	486	14	AA962900	L30-430T3 Ice plant La	3.12e-04
C	7	95	95.0	487	14	AA962902	L30-433T3 Ice plant La	3.12e-04
C	8	85	85.0	209	8	AA213399	z191g10.r1 NCI-CGAP_GC	9.90e-02
C	9	85	85.0	236	16	N41298	YW68604.r1 Homo sapien	9.90e-02
C	10	85	85.0	285	16	T35252	EST81954 Homo sapien	9.90e-02
C	11	85	85.0	290	19	H92421	YS86603.r1 Homo sapien	9.90e-02
C	12	85	85.0	293	21	T19059	B060055 Testis 2 Homo	9.90e-02

C	13	85	85.0	305	16	T34361	EST66604 Homo sapien	9.90e-02
C	14	85	85.0	371	26	B45681	HS-1062-A2-A06-MF.ab1	9.90e-02
C	15	85	85.0	377	15	C83092	Oryctolagus cuniculus	9.90e-02
C	16	85	85.0	377	15	C83225	Oryctolagus cuniculus	9.90e-02
C	17	85	85.0	380	24	AA295789	EST100992 Pancreas tum	9.90e-02
C	18	85	85.0	390	5	T99944	ye72e03.r1 Homo sapien	9.90e-02
C	19	85	85.0	416	5	T75533	ye63e02.r1 Homo sapien	9.90e-02
C	20	85	85.0	420	5	T99283	ye63f01.r1 Homo sapien	9.90e-02
C	21	85	85.0	443	14	AA837379	od41c11.s1 NCI-CGAP_GC	9.90e-02
C	22	85	85.0	482	24	AA313767	EST185694 Colon carcin	9.90e-02
C	23	85	85.0	504	6	AA195688	z132h11.r1 Soares 2NM	9.90e-02
C	24	80	80.0	169	18	AA159267	v288e11.r1 Soares 2NM	1.46e+00
C	25	80	80.0	307	15	AA966374	OS06c05.s1 NCI-CGAP_GC	1.46e+00
C	26	80	80.0	310	19	HSC2SE022	H. sapiens partial cDN	1.46e+00
C	27	80	80.0	321	9	AA597789	29856 Lambda-PRL2 Arab	1.46e+00
C	28	80	80.0	334	16	H00125	Y169405.r1 Homo sapien	1.46e+00
C	29	80	80.0	351	12	AA731954	nw64h07.s1 NCI-CGAP_GC	1.46e+00
C	30	80	80.0	364	16	HSC16A051	H. sapiens partial cDN	1.46e+00
C	31	80	80.0	390	13	AA810584	oa85h10.s1 NCI-CGAP_GC	1.46e+00
C	32	80	80.0	406	7	AA447072	z288602.r1 Soares tota	1.46e+00
C	33	80	80.0	407	26	B63132	CIT978SK-A-657C7.rp CI	1.46e+00
C	34	80	80.0	413	13	AA180411	z232p03.r1 Homo sapien	1.46e+00
C	35	80	80.0	414	20	N99782	Y199f01.r1 Homo sapien	1.46e+00
C	36	80	80.0	421	16	H01380	HS-3218.B1.B03.MR CIT	1.46e+00
C	37	80	80.0	429	28	AO181628	HS-1056-B1-C03-MF.ab1	1.46e+00
C	38	80	80.0	440	26	B42802	z248c08.s1 NCI-CGAP_GC	1.46e+00
C	39	80	80.0	467	8	AA285125	Y11b10.r1 Homo sapien	1.46e+00
C	40	80	80.0	484	16	R62848	aag3e09.s1 Homo sapien	1.46e+00
C	41	80	80.0	489	13	AA464930	v978b09.r1 Barstead MP	1.46e+00
C	42	80	80.0	491	7	AA473537	aj44b12.s1 Soares test	1.46e+00
C	43	80	80.0	545	12	AA813019	CIT-HSP-236108.r1 CIT	1.46e+00
C	44	80	80.0	562	28	AO111189	T26N18-r7 TAMU Arabido	1.46e+00
C	45	80	80.0	1005	26	B11855		

ALIGNMENTS

RESULT 1
 LOCUS A1007467 402 bp mRNA EST 12-JUN-1998
 DEFINITION L0-370M13R Ice plant Lambda Uni-Zap XR expression library, 0 hours
 NACL treatment Mesembryanthemum crystallinum cDNA clone L0-370 5'
 similar to 40S RIBOSOMAL PROTEIN S4, mRNA sequence.

ACCESSION A1007467
 NID 93216977
 KEYWORDS

SOURCE
 ORGANISM Mesembryanthemum crystallinum
 Common ice plant.

REFERENCE
 1 (bases 1 to 402)
 Cushman, J.C.
 AUTHOR
 TITLE An expressed sequence tag database for the common ice plant,
 Mesembryanthemum crystallinum
 JOURNAL
 COMMENT Unpublished (1997)

JOURNAL
 COMMENT

Contact: Cushman JC
 Department of Biochemistry and Molecular Biology
 Oklahoma State University
 350 Noble Research Center, Stillwater, OK 74078-3035, USA
 Tel: 405-744-6207
 Fax: 405-744-7799
 Email: jcushman@biochem.okstate.edu

PCR Primers
 FORWARD: T7
 BACKWARD: T3

Seq primer: M13 reverse
 High quality sequence stop: 300.
 Location/Qualifiers

FEATURES
 1. 402
 /organism="Mesembryanthemum crystallinum"

```

Query Match      95.0%  Score 95;  DB 15;  Length 402;
Best Local Similarity 63.6%  Pred. No. 3.12e-04;
Matches 14;  Conservative 5;  Mismatches 3;  Indels 0;  Gaps 0;

BASE COUNT      110 a      84 c      93 g      115 t
ORIGIN
      20 CATGCCATTCTGATGACACT 41
      |||||:||||:||||:|
      22 CTTTCCRTTGTGTGARNRGT 1

RESULT 2
LOCUS      AA962868      431 bp      mRNA      EST      18-MAY-1998
DEFINITION L30-398r3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-398 5'
            similar to 40S ribosomal protein S4, type I, mRNA sequence.
ACCESSION  AA962868
NID         93136341
KEYWORDS   EST.
SOURCE     Mesembryanthemum crystallinum
ORGANISM   Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Caryophyllidae; Caryophyllales; Alstroceae;
            Mesembryanthemum.
REFERENCE   1 (bases 1 to 431)
AUTHORS   Cushman, J.C.
TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
COMMENT   Unpublished (1997)

JOURNAL
COMMENT   Contact: Cushman JC
            Department of Biochemistry and Molecular Biology
            Oklahoma State University
            350 Noble Research Center, Stillwater, OK 74078-3035, USA
            Tel: 405-744-6207
            Fax: 405-744-7799
            Email: jcushman@biochem.okstate.edu
            PCR Primers
            FORWARD: T7
            BACKWARD: T3
            Plate: L30-5
            Seq primer: T3
            row: A
            column: 2
            High quality sequence stop: 310.
            Location/Qualifiers
                1..431
                /organism="Mesembryanthemum crystallinum"
                /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                EcoRI; Site_2: XhoI"
                /db_xref="taxon:3544"
                /db_xref="taxon:3544"
                /clone="L30-398"
                /clone_lib="Ice plant Lambda Uni-Zap XR expression
                library, 30 hours NaCl treatment"
                /tissue_type="leaf"
                /dev_stage="Six week old"
                /dev_stage="Six week old"
BASE COUNT      110 a      84 c      93 g      125 t
ORIGIN
      210 CATGCCATTCTGATGACACT 231
      |||||:||||:||||:|

```

```

CP      22  CRTTNCRTTYTGRTGNARNGT  1

RESULT      3
LOCUS      AA962918      471 bp      mRNA      EST      18-MAY-1998
DEFINITION L30-449T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-449 5
ACCESSION  AA962918
NID
KEYWORDS   Similar to 40S ribosomal protein S4 type I, mRNA sequence.
SOURCE     93136391
          EST.
          common ice plant.
          Mesembryanthemum crystallinum
          Eukaryota; Viridiplantae; Caryophyta/Embryophyta group;
          Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
          Eudicotyledons; Caryophyllidae; Caryophyllales; Alzooceae;
          Mesembryanthemum.
          1 (bases 1 to 471)
          Cushman, J.C.
          An expressed sequence tag database for the common ice plant,
          Mesembryanthemum crystallinum
          Unpublished (1997)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Cushman JC
Department of Biochemistry and Molecular Biology
Oklahoma State University
350 Noble Research Center, Stillwater, OK 74078-3035, USA
Tel: 405-744-6207
Fax: 405-744-7799
Email: jcushman@biochem.okstate.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L30-5 row: E column: 5
Seq primer: T3
High quality sequence stop: 300.
Location/Qualifiers
1..471
/organism="Mesembryanthemum crystallinum"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site1:
EcoRI; Site2: XhoI"
/db_xref="taxon:3544"
/clone="L30-449"
/clone_lib="Ice plant Lambda Uni-Zap XR expression
library, 30 hours NaCl treatment"
/tissue_type="leaf, 30 h 0.4M NaCl"
/dev_stage="Six week old"
111 c 101 g 134 t

BASE COUNT      125 a
ORIGIN           111 c 101 g 134 t

Query Match      95.0%; Score 95; DB 14; Length 471;
Best Local Similarity 63.6%; Pred. No. 3,12e-04;
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db      210  CATGCCATCTCATGCAACGT 231
|:|:|:|:|:|:|:|:|:|:|:|:|
CP      22  CRTTNCRTTYTGRTGNARNGT  1

RESULT      4
LOCUS      AA887369      472 bp      mRNA      EST      30-MAR-1998
DEFINITION L30-342T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-342 5,
            similar to 40S ribosomal protein S4, type I, mRNA sequence.
ACCESSION  AA887369
NID
KEYWORDS   common ice plant.
          Mesembryanthemum crystallinum
          Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
          Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
          Eudicotyledons; Caryophyllidae; Caryophyllales; Alzooceae;
          Mesembryanthemum.
          1 (bases 1 to 472)
          Cushman, J.C.
          An expressed sequence tag database for the common ice plant,
          Mesembryanthemum crystallinum
          Unpublished (1997)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Cushman JC
Department of Biochemistry and Molecular Biology
Oklahoma State University
350 Noble Research Center, Stillwater, OK 74078-3035, USA
Tel: 405-744-6207
Fax: 405-744-7799
Email: jcushman@biochem.okstate.edu
PCR Primers
FORWARD: T7
BACKWARD: T3
Plate: L30-5 row: E column: 5
Seq primer: T3
High quality sequence stop: 300.
Location/Qualifiers
1..472
/organism="Mesembryanthemum crystallinum"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site1:
EcoRI; Site2: XhoI"
/db_xref="taxon:3544"
/clone="L30-342"
/clone_lib="Ice plant Lambda Uni-Zap XR expression
library, 30 hours NaCl treatment"
/tissue_type="leaf, 30 h 0.4M NaCl"
/dev_stage="Six week old"
111 c 101 g 134 t

```

Query Match 95.08; Score 95; DB 14; Length 486;

Eukaryotae: M

REFERENCE

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 305)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,

Bult, C. J., Lee, N., Kihnres, E. F., Weinstock, K. G., Gocayne, J. D., White, O., Sutton, G., Blake, J. A., Brandon, R. C., Chiu, M.-W., Clayton, R. A., Cline, R. T., Cotton, M. D., Earle-Hughes, J., Fine, L. D., Fitzgerald, L. M., Fitzhugh, W. M., Friedman, J. L., Hinkle, N. S. M., Glodde, A., Gnehm, C. L., Hanna, M. C., Hedblom, E., Hoque Jr, P. S., Kelley, J. M., Klinek, K. M., Kelley, J. C., Liu, L.-I., Marmaros, S. M., Merrick, J. M., Moreno-Palauques, R. F., McDonald, L. A., Nguyen, D. T., Pauleggi, S. M., Phillips, C. A., Ryder, S. E., Scott, J. L., Seeger, D. M., Shirley, R., Small, K. V., Srinivas, T. A., Tienbach, T. P.

TITTLE
JOURNAL

COMMENT

The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tshih@tigr.org

For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@db.tigr.org).

1 Location/Qualifiers

305

SOA

MRNA

Query Match	85.0%	Score 85;	DB 16;	Length 305;
Best Local Similarity	59.1%	Pred. No. 9,90e-02;		
Matches 13;	Conservative 5;	Mismatches 4;	Indels 0;	Gaps 0;
Db	29	CCTTCACACAGCAGCGAAGCT	50	
	: : : :			

LOCUS	371 bp	DNA	GSS	20-OCT-1997
DEFINITION	B45681			
	HS-1062-A2-A06-MF.abi	CIT Human Genomic Sperm Library C Homo		

ACCESSION
NID
KEYWORDS

SOURCE ORGANISM

AUTHORS

TITLE

**JOURNAL
COMMENT**

Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 305)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,

TITLE	Initial Assessment of Human Gene Diversity and Expression Pattern
JOURNAL	Based Upon 52 Million Basepairs of cDNA Sequence
COMMENT	Unpublished (1995)
	Other_ESTS: THC9497

The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tshirif@tigr.org

For clone availability, additional sequence and expression information related to this EST, please contact the TRIC Database (tdbinfo@tdb.tigr.org).

```

FEATURES
SOURCE      : 1 305
              location/qualifiers

```

ORIGIN	BASE COUNT	MRNA	ORIGANISM="Homo sapiens"
75 a	84 c	92 g	52 t
2 others			

Query Match	85.08;	Score 85;	DB 16;	Length 305;
Best Local Similarity	59.18;	Pred. No. 9.90e-02;		
Matches	13;	Conservative	5;	Mismatches 4;
			Indels	0;
			Gaps	0;

```

Db      29 CCTTCACCGAGCATGGCAACGT 50
      | : | | | : | : | | | : |
Qy      2 CNYTNCAVCARAAYGGNAAYGT 23

```

Accession	LOCUS	14				
	B45681	371 bp	DNA	GSS	20-OCT-1997	
	HS-1062-A2-A06-MF.abi			CIT Human Genomic Sperm Library C Homo		
	sapiens genomic clone			Plate=Ct 764 Col=12 Row=A, genomic survey		
	sequence.					

ACCESSION
NID
KEYWORDS

SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria
Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 371)

AUTHORS Mairiaux, G.G., Zackrore, K.D., Smith, T., Tipton, S., Schmidt, S.,
Tritcoff, R., Abbajian, C., Blanchard, A., West, A. and Hood, L.E.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequences
Tagged Connectors
JOURNAL Unpublished (1997)

Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington

2/

THIS PAGE BLANK (USPTO)

Release 3.1A John F. Collins, Blocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MSPRCH_gp  protein - protein database search, using Smith-Waterman algorithm

Run on:      Fri Apr 16 14:18:00 1999;  MasPar time 3.81 Seconds
Tabular output not generated.  167.077 Million cell updates/sec

```

Title:	>US-09-011-797-2
Description:	(1-17) from US09011797.pep
Perfect Score:	116
Sequence:	1 FGGFTGARKSARKLANQ 17

Scoring table: PAM 150
Gap 15

Searched: 116695 seqs, 37453910 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: plr58

Statistics: Mean 26.726; Variance 36.801; scale 0.726

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	116	100.0	109	2	S70631	orphanin FQ/nocicept	5.09e-133
2	116	100.0	168	2	S60337	opioid receptor-like	5.09e-133
3	116	100.0	176	2	JC6152	orphanin FQ precursor	5.09e-133
4	116	100.0	181	2	JC6151	orphanin FQ precursor	5.09e-133
5	116	100.0	187	2	JC4502	opioid-receptor like	5.09e-133
6	67	57.8	205	2	S73721	ribosomal protein S4	7.35e-020
7	66	56.9	205	2	D64334	ribosomal protein S4	1.16e-011
8	66	56.9	526	2	S74025	conserved hypothetical	1.16e-011
9	61	52.6	405	1	DEPMNA	glyceraldehyde-3-phos	1.08e+000
10	60	51.7	310	2	A55768	asparaginyl-peptide a	1.66e+000
11	59	50.9	177	2	C69476	hypothetical protein	2.55e+000
12	59	50.9	211	2	A70307	ribosomal protein S04	2.55e+000
13	59	50.9	231	2	C71532	probable ribonuclease	5.19e+000
14	58	50.0	208	2	F64681	ribosomal protein S4	3.89e+000
15	58	50.0	421	2	C69665	nitrate transporter n	3.89e+000
16	58	50.0	998	2	A36368	transcription factor	5.91e+000
17	57	49.1	159	2	H65167	hypothetical 18.2 kd	5.91e+000
18	57	49.1	528	2	B64760	hypothetical protein	5.91e+000
19	57	49.1	806	2	PC6010	RNA helicase Gu - hum	5.91e+000
20	56	48.3	301	2	S39151	csk1 protein - flsio	8.93e+000
21	55	47.4	173	2	S42246	NADH dehydrogenase (u	1.34e+001
22	55	47.4	264	2	B64714	ribosomal protein S2	1.34e+001
23	55	47.4	470	2	S68964	legumin precursor (cl	1.34e+001

5	52	44.8	305	2	AA4362	NADH dehydrogenase (u	4.41e+01
6	52	44.8	125	2	S50170	mercury resistance op	4.41e+01
7	52	44.8	90	1	DNZRHM	DNA-binding protein H	4.41e+01
8	53	45.7	856	2	A64699	hypothetical protein	2.99e+01
9	41	45.7	704	2	S64000	probable membrane pro	2.99e+01
10	42	45.7	542	2	H69961	transcription regulat	2.99e+01
11	39	45.7	697	2	A64878	peptide transport per	2.99e+01
12	40	45.7	592	2	A64878	transcription regulat	2.99e+01
13	38	45.7	482	2	S31478	alpha-amylase (EC 3.2	2.99e+01
14	37	45.7	475	1	WNAAP6	early E1B 53K protein	2.99e+01
15	36	45.7	400	2	A71490	probable aspartic aa	2.99e+01
16	53	45.7	370	2	A71490	probable aspartic aa	2.99e+01
17	53	45.7	352	2	A69103	conserved hypothetical	2.99e+01
18	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
19	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
20	32	45.7	323	2	S64246	aspartate-semialdehyd	2.99e+01
21	32	45.7	345	2	E07094	probable asid protein	2.99e+01
22	53	45.7	352	2	A69103	conserved hypothetical	2.99e+01
23	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
24	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
25	29	45.7	141	2	S73848	ribosomal protein S19	2.99e+01
26	28	46.6	651	2	D70789	probable acetyl-coenz	2.01e+01
27	54	46.6	475	2	S73766	MG294 homolog A05_orf	2.01e+01
28	54	46.6	374	2	S61178	heat shock protein dn	2.01e+01
29	55	47.4	476	2	S68965	legumin precursor (cl	1.34e+01
30	25	46.6	224	2	F69091	ribosomal protein L10	2.01e+01
31	25	46.6	224	2	F69091	ribosomal protein L10	2.01e+01
32	26	46.6	374	2	S61178	heat shock protein dn	2.01e+01
33	27	46.6	475	2	S73766	MG294 homolog A05_orf	2.01e+01
34	28	46.6	651	2	D70789	probable acetyl-coenz	2.01e+01
35	29	45.7	93	2	S25976	ribosomal protein S19	2.99e+01
36	53	45.7	300	2	A42204	phosphoenolpyruvate m	2.99e+01
37	53	45.7	300	2	A42204	phosphoenolpyruvate m	2.99e+01
38	53	45.7	323	2	S64246	aspartate-semialdehyd	2.99e+01
39	53	45.7	345	2	E07094	probable asid protein	2.99e+01
40	53	45.7	352	2	A69103	conserved hypothetical	2.99e+01
41	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
42	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
43	32	45.7	323	2	S64246	aspartate-semialdehyd	2.99e+01
44	32	45.7	345	2	E07094	probable asid protein	2.99e+01
45	53	45.7	352	2	A69103	conserved hypothetical	2.99e+01
46	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
47	34	45.7	378	1	A95682	alcohol dehydrogenase	2.99e+01
48	29	45.7	141	2	S73848	ribosomal protein S19	2.99e+01
49	28	46.6	651	2	D70789	probable acetyl-coenz	2.01e+01
50	54	46.6	475	2	S73766	MG294 homolog A05_orf	2.01e+01
51	54	46.6	374	2	S61178	heat shock protein dn	2.01e+01
52	55	47.4	476	2	S68965	legumin precursor (cl	1.34e+01
53	25	46.6	224	2	F69091	ribosomal protein L10	2.01e+01
54	25	46.6	224	2	F69091	ribosomal protein L10	2.01e+01
55	26	46.6	374	2	S61178	heat shock protein dn	2.01e+01
56	27	46.6	475	2	S73766	MG294 homolog A05_orf	2.01e+01
57	28	46.6	651	2	D70789	probable acetyl-coenz	2.01e+01
58	29	45.7	93	2	S25976	ribosomal protein S19	2.99e+01

ALIGNMENTS

```

RESULT      1
ENTRY
TITLE      orphanin FQ/nocticeptin - mouse (fragment)
ORGANISM   #formal_name Mus musculus #common_name house mouse
DATE       14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997

ACCESSIONS
REFERENCE   S70631
#authors   Pan, Y.X.; Xu, J.; Pasternak, G.W.
#journal   Biochem. J. (1996) 315:11-13
#title     Cloning and expression of a cDNA encoding a mouse brain
           orphanin FQ/nocticeptin precursor.
#accession S70631
#status    preliminary
#molecule_type mRNA
#residues  1-109 #label PAN
#crossref  references EMBL:U44027; NID:q1335869; PID:q1335870
SUMMARY    #length 109 #checksum 9509

Query Match      100.0%; Score 116; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 5,09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0

DB      63 FGGFTGARKSARKLANQ 79
        |||||||||||||||
        1 FGGFTGARKSARKLANQ 17

RESULT      2
ENTRY
TITLE      S60337 #type fragment
           opiod receptor-like ORL(1) receptor antagonist - rat
           (fragment)
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change
18-Mar-1997

ACCESSIONS
REFERENCE   S60337
#authors   Weunier, J.C.; Mollereau, C.; Toll, L.; Snaudeau, C.;
           Molstead, C.; Alviñerie, P.; Butour, J.L.; Guillemot, J.C.;
           Ferrara, P.; Monsarrat, B.; Mazarguil, H.; Vassart, G.;
           Parnientier, M.; Costentin, J.
#journal   Nature (1995) 377:532-535
#title     Isolation and structure of the endogenous agonist of opiodoid
           receptor-like ORL(1) receptor.
#accession S60337
#status    preliminary; not compared with conceptual translation
#molecule_type mRNA

```

##residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 116; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 122 FGGFTGARKSARKLANQ 138
|||||
QY 1 FGGFTGARKSARKLANQ 17

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
JC6152
ACCESSIONS JC6151
REFERENCE Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6152
#molecule_type mRNA
#residues 1-176 ##label NOT
#cross-references GB:U48263; NID:q1185009; PID:q1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opiod receptors. It functions in development.
KEYWORDS opiod peptide
SUMMARY #length 176 #molecular-weight 20295 #checksum 498

Query Match 100.0%; Score 116; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 130 FGGFTGARKSARKLANQ 146
|||||
QY 1 FGGFTGARKSARKLANQ 17

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
JC6151
ACCESSIONS JC6151
REFERENCE Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6151
#molecule_type mRNA
#residues 1-181 ##label NOT
#cross-references GB:U48262; NID:q1185011; PID:q1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opiod receptors. It is involved in development.
KEYWORDS opiod peptide
SUMMARY #length 181 #molecular-weight 20478 #checksum 3303

Query Match 100.0%; Score 116; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 135 FGGFTGARKSARKLANQ 151
|||||
QY 1 FGGFTGARKSARKLANQ 17

RESULT 5
ENTRY JC4502 #type complete
TITLE opiod-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N23K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 10-Sep-1997
JC4502
ACCESSIONS JC4502
REFERENCE Salto, Y.; Maruyama, K.; Saldo, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label SAI
#cross-references DDBJ:DB2056; NID:q1857023; PID:di009395; PID:q1217906
REFERENCE JC4652
#journal Houtani, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto, T.
#title Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ precursor.
#accession JC4652
#molecule_type mRNA
#residues 1-187 ##label HOU
#cross-references DDBJ:DB2866; NID:q1311472; PID:di012281; PID:q1311473
#experimental_source brain
#note The authors translated the codon CTC for residue 72 as Lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions as a neuropeptide precursor and plays a role in neuronal differentiation and development.
GENETICS N23K
#gene brain; differentiation; G protein-coupled receptor; neuropeptide
FEATURES
1-187
109-126 #domain signal sequence #status predicted #label SIG
141-157 #product neuropeptide precursor, N23K #status predicted
#label MAY
#region D-A-E-P-G-A motif
160-176 #product neuropeptide #status predicted #label NRP
#product heptadecapeptide #status predicted #label HPP
SUMMARY #length 187 #molecular-weight 20884 #checksum 8758

Query Match 100.0%; Score 116; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 5.09e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 141 FGGFTGARKSARKLANQ 157
|||||
QY 1 FGGFTGARKSARKLANQ 17

RESULT 6
ENTRY S73721 #type complete
TITLE ribosomal protein S4 - Mycoplasma pneumoniae (ATCC 29342)
ALTERNATE_NAMES (SGC3)
ORGANISM hypothetical protein H08_orf205
#formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change

21-Aug-1998
ACCESSIONS S73721
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
#cross-references EMBL:9710585
#accession S73721
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-205 ##label HIM
##cross-references EMBL:AE000038; GB:U00089; NID:g1674074; PID:g1674077
#note the nucleotide sequence was submitted to the EMBL Data Library, November 1996

GENETICS
#gene rpsD
#genetic_code SGC3
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S4
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 205 #molecular-weight 23774 #checksum 29

Query Match
Best Local Similarity 57.8%; Score 67; DB 2; Length 205;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPTRSARQNVN 119
||:|||||:|
QY 3 GFTGARKSARKLANO 17

RESULT 7
ENTRY D64234
#type complete
TITLE ribosomal protein S4 - Mycoplasma genitalium (SGC3)
#formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 21-Aug-1998

ACCESSIONS D64234
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bort, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references EMBL:96026346
#accession D64234
#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues 1-205 ##label TIGR
##cross-references GB:U39712; GB:U43967; NID:g1046005; PID:g1046011; TIGR:MG311
##experimental_source strain G-37

GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S4
SUMMARY #length 205 #molecular-weight 23947 #checksum 310

Query Match
Best Local Similarity 56.9%; Score 66; DB 2; Length 205;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 105 GFAPTRSARQNVN 119
||:|||||:|
QY 3 GFTGARKSARKLANO 17

RESULT 8
ENTRY S74025
#type complete
TITLE conserved hypothetical protein c0207 - Sulfolobus solfataricus
ORGANISM #formal_name Sulfolobus solfataricus
DATE 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 21-Aug-1998

ACCESSIONS S74025
REFERENCE S73076
#authors Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.L.; Young, F.; Schenk, M.E.; Gaasterland, T.; Doolittle, W.F.; Ragan, M.A.; Charlebois, R.L.
#journal Mol. Microbiol. (1996) 22:175-191
#title Organizational characteristics and information content of an archaeal genome: 156 kb of sequence from Sulfolobus solfataricus P2.

#accession S74025
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-526 ##label SEN
##cross-references EMBL:Y08256; NID:g1707679; PID:e283902; PID:g1707719
#experimental_source strain P2
#note the nucleotide sequence was submitted to the EMBL Data Library, September 1996

CLASSIFICATION #superfamily proline carrier protein
SUMMARY #length 526 #molecular-weight 56210 #checksum 8544

Query Match
Best Local Similarity 56.9%; Score 66; DB 2; Length 526;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 220 FGGFSGAFHNAQTLSQ 236
|||||:|:|:|
QY 1 FGGFSGARKSARKLANO 17

RESULT 9
ENTRY DEPMNA
#type complete
TITLE glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) (EC 1.2.1.13) A precursor, chloroplast - garden pea
ORGANISM #formal_name Pisum sativum #common_name garden pea
DATE 05-Sep-1997
#cross-references EMBL:U08256; NID:g1707679; PID:e283902; PID:g1707719

ACCESSIONS S14243; S05554; A38270
REFERENCE A38270
#authors Lland, M.F.; Zhang, D.X.; Cerff, R.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8918-8922
#title Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus.
#cross-references EMBL:91062390
#accession S14243
##molecule_type DNA
##residues 1-405 ##label LIA
##cross-references EMBL:X52148; NID:g12158; PID:g12159
#journal S05552
#authors Brinkmann, H.; Cerff, R.; Salomon, M.; Soll, J.
#journal Plant Mol. Biol. (1989) 13:81-94
#title Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits gapA and gapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach.

#accession S05554
#molecule_type mRNA
#residues 1-405 ##label BRI
##cross-references EMBL:X15190
#note the authors translated the codon AGA for residue 194 as Gly; the sequence shown follows the authors' translation

GENETICS
#gene Gpal

```

#journal
#title
The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession
C69476
#status
preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-177 ##label KLE
#cross-references GB:AE000978; GB:AE000782; NID:g2689301; PID:g2648748;
TIGR:AF1812
SUMMARY
#length 177 #molecular-weight 20596 #checksum 5189

Query Match 50.9%; Score 59; DB 2; Length 177;
Best Local Similarity 46.7%; Pred. No. 2.55e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 138 FLSYGSIRKARKLG 152
|:::|::|::|:
QY 1 FGGFGARKSARKIA 15

RESULT 12
ENTRY A70307 #type complete
TITLE ribosomal protein S04 - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
02-Jul-1998

ACCESSIONS
A70307
REFERENCE
A70300
#authors
Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
Nature (1998) 392:353-358
The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession
A70307
#status
preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-211 ##label AQP
#cross-references GB:AE000672; NID:g2982810; PID:g2982819; GB:AE000657
#experimental_source strain VFS

GENETICS
#gene
rpsD
CLASSIFICATION
#superfamily Escherichia coli ribosomal protein S4
SUMMARY
#length 211 #molecular-weight 24815 #checksum 7435

Query Match 50.9%; Score 59; DB 2; Length 211;
Best Local Similarity 40.0%; Pred. No. 2.55e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 108 GFASRRQARQIVAH 122
|:::|::|::|:
QY 3 GFTGARKSARKLANQ 17

RESULT 13
ENTRY C71532 #type complete
TITLE Probable ribonuclease III - Chlamydia trachomatis (serotype
D, strain WM3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
13-Sep-1998
C71532
A71460
ACCESSIONS
A71460
REFERENCE
#authors
Stephens, R.S.; Kalman, S.; Lammel, C.U.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, O.; Koonin, E.V.; Davis, B.W.

```



```

#submision      submitted to GenBank, May 1998
#description    Genome sequence of an obligate intracellular pathogen of
#               humans: Chlamydia trachomatis.
#accession      C71532
#status         Preliminary
##molecule_type DNA
##residues      1-231 ##label ARN
##cross-references GB:AE001302; GB:AE001273; NID:g3328708; PID:g3328712
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene
#inc
SUMMARY
#length 231 #molecular-weight 25561 #checksum 9834
Query Match      50.9%; Score 59; DB 2; Length 231;
Best Local Similarity 61.5%; Pred. No. 2.55e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 207 GFAGKKEAKLA 219
|||:| | | |
3 GFGARKSARKLA 15

RESULT 14
ENTRY      F64681 #type complete
TITLE      ribosomal protein S4 - Helicobacter pylori (strain 26695)
ORGANISM   #formal_name Helicobacter pylori
DATE       09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
31-Oct-1997
ACCESSIONS F64681
REFERENCE   A64520
#authors    Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
            Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
            H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
            J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
            Richardson, D.; Dodson, R.; Khairak, H.G.; Glodek, A.;
            McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
            Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
            Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
            Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
            W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
            C.M.; Venter, J.C.
            Nature (1997) 388:539-547
            The complete genome sequence of the gastric pathogen
            Helicobacter pylori.
            #cross-references M01D:97394467
            #accession F64681
            #status preliminary; nucleic acid sequence not shown;
            translation not shown
##molecule_type DNA
##residues      1-208 ##label TOM
##cross-references GB:AE000633; GB:AE000511; NID:g2314452; PID:g2314460;
            TIGR:HP1294
CLASSIFICATION #superfamily Escherichia coli ribosomal protein S4
SUMMARY      #length 208 #molecular-weight 23964 #checksum 6652
Query Match      50.0%; Score 58; DB 2; Length 208;
Best Local Similarity 46.7%; Pred. No. 3.89e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 108 GFATRSSAROLVTH 122
||:| | | | |
3 GFGARKSARKLANO 17

RESULT 15
ENTRY      C69665 #type complete
TITLE      nitrate transporter nasa - Bacillus subtilis
ORGANISM   #formal_name Bacillus subtilis
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS C69665; I40026
REFERENCE   A69580
#authors    Kust, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

```

```

Alloni, G.; Azavedo, V.; Berto, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Bourisier, L.; Brans,
A.; Braun, M.; Brigneau, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codan, J.F.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haeche, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Katamata, D.;
Kasahara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konungstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Paro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pujić, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Sero, S.J.; Serror, P.; Shin, B.S.; Solido,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references M01D:98044033
#accession C69665
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues      1-421 ##label KUN
##cross-references GB:Z99105; GB:AL009126; NID:g2632457; PID:e1182285;
            PID:g2632619
            #experimental_source strain 168
REFERENCE   I39887
#authors    Ogasawara, N.; Fujita, Y.; Kobayashi, Y.; Sadate, Y.; Tanaka,
            T.; Takahashi, H.; Yamane, K.; Yoshikawa, H.
            Microbiology (1995) 141:257-259
            Systematic sequencing of the Bacillus subtilis genome:
            progress report of the Japanese group.
            #cross-references M01D:95215077
            #accession I40026
            #status translated from GB/EMBL/DBJ
            ##molecule_type DNA
            ##residues      'U', 1-421 ##label RES
            ##cross-references GB:D30689; NID:g710016; PID:g710017
GENETICS
#gene      nasa
#inc
SUMMARY      #length 421 #molecular-weight 46067 #checksum 556
Query Match      50.0%; Score 58; DB 2; Length 421;
Best Local Similarity 61.5%; Pred. No. 3.89e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 GKGAKARKARRM 400
||:| | | | |
2 GFGARKSARKLA 14

```

Wed Apr 28 09:08:33 1999

US-09-011-797-2.rpr

 WIRELESS
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Search: n.a. - n.a. Smith-Waterman search, using a protein query
 which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:35:11 1999; Master time 22.17 Seconds
 Tabular output not generated. 147.283 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect score: 100
 N.A. Sequence: 1 ACNTYNCAYCARAAVGNAAVGTN 24
 Comp: TGNRANGRTGYTTRCNCNTTCAN

Scoring table: TABLE backtranslated
 Gap 40

Mismatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-gene32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40

Statistics: Mean 30.944; Variance 103.177; scale 0.300
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	100	100.0	932 34	T79888	DNA encoding human op	1.11e+00
2	85	85.0	2158 34	O14624	Plasmid pPRDPPD inser	2.77e+01
3	85	85.0	2158 21	T34372	Plasmid pPRDPPD (ATCC	2.77e+01
4	80	80.0	857 1	N93084	Lambda gt 11 pL1371	7.71e+01
5	80	80.0	1763 3	O14071	KpnI-XbaI fragment fr	7.71e+01
6	80	80.0	1763 1	O03742	Gene encodes function	7.71e+01
7	80	80.0	2839 1	O03742	Human SKI related gen	7.71e+01
8	80	80.0	2888 1	O03743	Human SKI related gen	7.71e+01
9	80	80.0	5687 29	T65001	Mouse cell cycle regu	7.71e+01
10	75	75.0	339 13	O68135	Sequence encoding hum	2.09e+02
11	75	75.0	339 13	O68132	Sequence encoding hum	2.09e+02
12	75	75.0	339 13	O68134	Sequence encoding hum	2.09e+02

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
13	75	75.0	339 13	O68135	Sequence encoding hum	2.09e+02
14	75	75.0	339 13	O68115	Sequence encoding hum	2.09e+02
15	75	75.0	339 13	O68146	Sequence encoding hum	2.09e+02
16	75	75.0	1315 30	T66684	Soybean peroxidase SE	2.09e+02
17	75	75.0	1455 17	T00777	Bacillus sp. alkaline	2.09e+02
18	75	75.0	1462 10	O66070	Sequence encoding soy	2.09e+02
19	75	75.0	2097 38	V01546	Corn starch branching	2.09e+02
20	75	75.0	2100 3	O13985	Rat synactin 1A gene	2.09e+02
21	75	75.0	2165 31	T69736	Gamma-cyclodextrin g1	2.09e+02
22	75	75.0	2235 1	N90389	Corn starch branching	2.09e+02
23	75	75.0	2467 15	O90271	DNA encoding polypept	2.09e+02
24	75	75.0	2665 31	T69729	Saccharomyces sp. rec	2.09e+02
25	75	75.0	2855 39	T97297	Plasmid pBE240 inser	2.09e+02
26	75	75.0	2855 39	V00011	Human sex comb on mid	2.09e+02
27	75	75.0	3065 39	V00014	Human sex comb on mid	2.09e+02
28	75	75.0	3065 39	V00014	Human sex comb on mid	2.09e+02
29	75	75.0	3065 39	V00010	Mouse sex comb on mid	2.09e+02
30	75	75.0	3097 31	T69590	Mouse sex comb on mid	2.09e+02
31	75	75.0	3255 39	V00013	Murine short form ob	2.09e+02
32	75	75.0	3255 39	V00013	Human sex comb on mid	2.09e+02
33	75	75.0	3327 39	V00012	Human sex comb on mid	2.09e+02
34	75	75.0	3327 39	T97298	Human sex comb on mid	2.09e+02
35	75	75.0	3337 1	N81089	Human sex comb on mid	2.09e+02
36	75	75.0	4335 37	T89387	Fragment of Pseudomon	2.09e+02
37	75	75.0	4355 2	O10947	Bovine coronavirus E2	2.09e+02
38	75	75.0	5255 3	O25448	Bovine Coronavirus E2	2.09e+02
39	75	75.0	5547 35	T68844	Photorehabus lunifesc	2.09e+02
40	75	75.0	6161 32	T77865	E.coli alkaline phosph	2.09e+02
41	75	75.0	7551 35	T68843	Photorehabus lunifesc	2.09e+02
42	75	75.0	8585 28	T48867	Brassica napus micros	2.09e+02
43	75	75.0	8585 1	O05749	Microspore-specific C	2.09e+02
44	75	75.0	580073 27	T58840	Mycoplasma genitalium	2.09e+02
45	70	70.0	4287 40	V20475	Human AML1/MTG8 Oncog	5.47e+02

ALIGNMENTS

RESULT 1
 ID T79888 standard; DNA; 932 BP.
 AC T79888;
 DT 08-DEC-1997 (first entry)
 DE DNA encoding human opiod receptor-like 1 receptor ligand.
 KW Opiod receptor-like 1; ORL1; ligand; pronociceptive; stress;
 KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
 KW memory; attention; sensory perception; learning; homeostasis;
 KW hypoaesthesia; nociceptin; endorphin; dynorphin A; ss.
 OS Synthelec.
 FH Key
 FT Location/Qualifiers
 FT cds 1..307
 FT /tag= a
 FT /note= "no start codon"
 MO9707208-A1.
 PN 27-FEB-1997.
 PD 14-AUG-1996; BE0087.
 PR 15-AUG-1995; US-002366.
 PA (ULBR) UNIV LIBRE BRUXELLES.
 PI Meunier J, Mollereau C, Parmentier M, Vassart G;
 DR P-PSDB; W25162, W25163, W25164.
 PT Novel ligand for the opiod receptor-like receptor, nociceptin -
 PT has pronociceptive properties, useful for treating or preventing
 PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
 PT etc.
 PS Claim 1: Page -; 48pp; English.
 CC T79888 encodes a ligand to human opiod receptor-like 1 (ORL1) receptor
 CC designated nociceptin, which resembles the endorphin dynorphin A.
 CC Inhibitors of nociceptin or its DNA, e.g. antisense sequences and
 CC antibodies can be used as new types of drugs in the control of various
 CC behaviours or functions. The inhibitors can be used to prevent or treat
 CC a disease related to hyperalgesia, neuroendocrine secretion, stress,
 CC locomotor activity, anxiety, instinctive behaviour, learning and memory,
 CC homeostasis, hyperalgesia, hypoaesthesia and/or sensory perception.
 CC Sequence 932 BP; 224 A; 273 C; 238 G; 197 T;

	Query Match	100.0%	Score 100;	DB 34;	Length 932;
	Best Local Similarity	65.2%;	Pred.	No. 1.11e+00;	
	Matches	15;	Conservative	5;	Mismatches 3; Indels 0; Gaps 0
D8	481 actctgcaccgaagatcgtcatct	503			
	: : : :				
OY	1 ACNNTMCACCAARAAGGNAATGT	23			

RESULT	ID	2	standard; CDNA; 2158 BP.
AC	014624;		
DT	30-JAN-1992	(first entry)	
DE	Plasmid pPAVPD	insert encoding a dunce-like phosphodiesterase.	
KW	rat brain; ATCC # 68586;	cyclic nucleotides; PDE; ss.	
RM	Rattus.		

FT	Key	Location/Qualifiers
FT	cds	1..1689
FT		/*tag- a
PD	MO9116457-A.	
PD	31-OCT-1991.	
PE	19-APR-1991.	U02714.
PR	20-APR-1990.	US-511715.
PA	(COLD-) COLD SPRING HARBOR.	
PI	Wigler MH, Colicelli JJ.	
DR	WPI: 91-339841/46.	
LR	P-PSDB: R14836.	
PT	Complementary screening for genes and prods. - e.g. RAS protein	
PT	and cAMP, that modify complement or suppress genetic defect and	
PT	correct associated phenotypic alteration	
PS	Claim 15, Page 58; 16pp; English.	
CC	Plasmid pRATDPD was isolated from a rat brain cDNA library. A large	
CC	open reading frame was found and a 562 amino acid sequence deduced	
CC	from it. The coding sequence is homologous to the Drosophila dance	
CC	gene; the proteins encoded by both sequences have 80 per cent identity	
CC	without the introduction of gaps, over a 252 amino acid region.	
CC	The dance gene encodes a cyclic nucleotide pde).	
CC	Sequence 2158 bp: 623	

Query Match	85.0%;	Score 85;	DB 3;	Length 2150;
Best Local Similarity	59.1%;	Pred. No. 2.77e+01;		
Matches	13; Conservative	5; Mismatches	4; Indels	0; Gaps 0;

```

Db      670 cattaccattctgatgtgcgt 691
      1:| | | | | | | | : | |
Cp      22 CRTTNCRTTYTGRTGNARNGT 1

```

RESULT	3
ID	T34372
NC	standard; CDNA; 2158 BP.
CC	U34372

09-OCT-1996 (first entry)
Plasmid PRATDPD (ATCC 68586) insert.
Rat brain derived.

KW heat shock sensitivity; PRAPPD: alternate splicing; rat DPD cDNA:
KW high affinity GAMP binding site
KW expression vector; PADNS; Ras(vall9); strain TK161-R2V; Drosophila;
KW heat shock sensitivity; PADNS; Ras(vall9); strain TK161-R2V; Drosophila;
KW expression vector; PADNS; Ras(vall9); strain TK161-R2V; Drosophila;
KW heat shock sensitivity; PADNS; Ras(vall9); strain TK161-R2V; Drosophila;
KW expression vector; PADNS; Ras(vall9); strain TK161-R2V; Drosophila;

OS	Rattus rattus.
EFH	Key
ET	cds
	Location/Qualifiers
	1-3701

US5527896-A. Dunce-like phosphodiesterase

18-JUN-1996.
20-APR-1990; 511715.
20-APR-1990; US-511715.

19-APR-1991; US-688352.
(COLD-) COLD SPRING HARBOR LAB.
Colicelli JT, Wigler MH;

WT: 96-299902/30.
P-PSDB; W00090.
DNA mols. isolated from human glioblastoma cells - encode

PT RAS-related or cyclic nucleotide phosphodiesterase proteins
PS Claim 4; Column 29-34; 101pp; English.
CC This sequence represents a dunce-like phosphodiesterase. The
CC derived cDNA inserts encoding a dunce-like phosphodiesterase. The
CC cDNA brain cDNA library was cloned into the yeast expression vector
CC PADMS. DNA prepared from the cDNA expression library was used to
CC transform the RAS(wall19) yeast strain YK101-R2V. Only one transformant
CC displayed heat shock sensitivity which is conditional on retention of
CC the expression plasmid. The plasmid designated PRATDP was isolated
CC from the transformant and sequenced. A large open reading frame of 562
CC codons was found. The first ATG appears at codon 46 and a protein which
CC initiates at this codon would have a predicted mol. wt. of approx. 60
CC kD. The isolated sequence was found to have homology to the *Drosophila*
CC dunce gene. The encoded proteins have an 806 amino acid identity
CC without the introduction of gaps over a 252 amino acid region located in
CC the centre of the rat *DPD* cDNA. The dunce gene has been shown to encode
CC a high affinity cAMP phosphodiesterase. The established open reading
CC frame in this sequence remains open for an additional 116 codons after
CC the termination codon, followed by more terminator codons, a
CC polyadenylation signal, and a polyadenine stretch. This 3' open reading
CC frame could be incorporated into another dunce-like phosphodiesterase
CC through alternate splicing.
SQ Sequence 2158 BP; 623 A; 535 C; 516 G; 100 T.

Query Match	85.08;	Score 85;	DB 21;	Length 2158;
Best Local Similarity	59.18;	Pred. NO. 2.77e+01;		
Matches	13;	Consensus 4;		

	conservative	5;	Mismatches	4;	Indels	0;	Gaps	0;
D670	cattaccatctgatgtgcgcg	691						

Cp 22 CRTTNCRTTYTGRTGNARNGT 1

	RESULT	4
ID	N93084	standard; cDNA; 857 BP.
AC	N93084;	

DT	29-JUN-1990	(first entry)
DE	Lambda gt 11 pl(13T)	contg. cDNA encoding Der p I antigen
KW	Der p I; house dust mite; allergy; asthma; ss.	

US	dermatophagoides pteronyssinus.
FH	Key
FT	Location/Qualifiers
cds	1..30

```

      /tag= a
      /product= prepro Der p 1
      31..735
      cds

```

```

/*lag= b
/product= pro Der p I
72..735
/****=

```

```
FT /lag= C
FT /product= mature Der p I
FT 803..808
FT /tag= C
```

PN	/	cyg -	C
WO8810297-A.			
29-DEC-1988.			
17-JUN-1988;			
PPF			AU0195.

PR
17-JUN-1987; AU-002523.
PA
(PRIN-) Princess Margaret Children's Med
Cancer Res, (HALL-) Walter and Eliza Hal

Thomas WR, Stewart GA, Turner KJ, Simpson WPI; 89-023841/03.
P-PSDB; P94864.

Cloning of mite allergens from dermatophagoides poultoni and dermatophagoides pteronyssinus polypeptide for use as diagnostic reagent substances for desensitisation.

Claim 3; Fig. 1; 35pp; English.
The cDNA encodes Der p 1, a 27 kD glycoprotein that elicits IgE anti-mite antibody causing asthma.

dermatitis. The clone, lambda gt 11 (13T) subcloned into vector pGEX and expressed glutathione transferase molecule. The polyclonal antibody was raised against the

diagnostic reagent since it reacts with IgE from serum of patients with allergy to mites. It can also be used for desensitisation treatment.

```

CC See also N93085.
S0 Sequence 857 BP: 294 A: 159 C: 154 G: 240 T:

Query Match
Best Local Similarity 57.1%; Score 80; DB 1; Length 857;
Matches 12; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 135 caccgccattcgtatgcgaag 155
      |||:||||:||||:|
Cp 22 CRTTNCRTTYTGRTGNARNG 2

RESULT 5
ID Q14071 standard; DNA; 1763 BP.
AC Q14071;
DT 10-JAN-1992 (first entry)
DE KpnI-BdaI fragment from pBY503 in Brevibacterium stationis IF012144.
KW PCR30; PCR30; genetic engineering; ss.
OS Brevibacterium stationis IFO 12144.
PN J03210184-A.
PD 13-SEP-1991.
PE 11-JAN-1990; 004212.
PR 11-JAN-1990; JP-004212.
PA (MTP ) MITSUBISHI PETROCH KK.
DR WPI: 91-314588/43.

New plasmid vector for breeding improved Corynebacterium - contains
P3 specified DNA regions for replicating proliferation of improved
PT Corynebacterium
PS Disclosure; Page 4; 13pp; Japanese.
CC This is the sequence of a preferred DNA region (A) containing a gene
CC capable of maintaining stable replicative proliferation of a plasmid
CC in Corynebacterium cells. Plasmids of the invention are composed of
CC such a region (A), along with a region (B) carrying the replicative
CC proliferation function of the plasmid in Corynebacterium and a
CC region (C) containing a drug resistance marker gene. The fragment is
CC excised from plasmid pBY503 which is retained in B.stationis IFO
CC 12144 (FERM BP-2515). A 2.1kb KpnI-EcoRI fragment from the same
CC source can be used for the same function. Region (B) is preferably
CC excised from the same source or from plasmid pBY502 retained in
CC B.flavum MJ233 (FERM BP-1497).
SQ Sequence 1763 BP: 433 A: 436 C: 445 G: 449 T:

Db 1327 acattgccgtttgatg 1343
      |||:||||:||||:|
Cp 23 ACRTTNCRTTYTGRTG 7

Query Match
Best Local Similarity 70.6%; Score 80; DB 3; Length 1763;
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1327 acattgccgtttgatg 1343
      |||:||||:||||:|
Cp 23 ACRTTNCRTTYTGRTG 7

RESULT 6
ID 003232 standard; RNA; 1763 BP.
AC 003232;
DT 12-JUL-1990 (first entry)
DE Gene encodes function of maintaining a plasmid stably in a coryne-form
DE bacteria of the genus Brevibacterium.
KW pBY503; cryptophan synthetase; coryneform; ds.
OS Brevibacterium stationis.
PN EP-352763-A.
PD 31-JAN-1990.
PE 26-JUL-1989; 113775.
PR 27-JUL-1988; JP-185428.
PR 08-SEP-1988; JP-223399.
PR 25-JAN-1989; JP-014098.
PA (MTP) Mitsubishi Petroch KK.
PI Kohama K, Kobayashi M, Kurusu Y, Yukawa H;
DR WPI: 90-031347/05.

New DNA fragments which stabilise plasmids in coryneform bacteria -
PT and derived recombinant vectors which are retained in cells
PT without selection pressure.
PS Claim 6; Page 23; 27pp; English.
CC Plasmid pBY503 contains DNA fragment present in Brevibacterium stationis

```

```

CC Ifo 12144, contains gene which stabilises plasmids in the bacteria,
CC allowing maintenance with no selection pressure.
CC Tryptophan synthetase gene with associated promoter/operator system has
CC been maintained within a stable plasmid associated with the gene.
CC Sequence 1763 BP; 433 A; 436 C; 445 G; 449 T;
SQ

Query Match      80.0%; Score 80; DB 1; Length 1763;
Best Local Similarity 70.6%; Pred. No. 7,71e+01;
Matches 12; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 1327 acctgcgcgtttgatg 1343
||:||||:||||:||||
Cc 23 ACRTTCRCRTTTCRTG 7

RESULT
7
ID 003742 standard; DNA; 2839 BP.
AC 003742;
DT 15-AUG-1989 (first entry)
DE Human SKI related gene
KW Human sloan kettering institute related gene; cancer; ss.
FW Key Location/Qualifiers
FT cds 709..2765
/*tag= a
/label=human SKI-related protein

FN J02053485-A.
PD 22-FEB-1990.
PR 19-AUG-1988; 205906.
PR 19-AUG-1988; JP-205906.
PA (RIKA) Rikagaku Kenkyusho .
DR WPI; 90-103117-A.
P-PSDB: R03664.
PT Human SKI related gene - coded by base sequence including 709 adenine to
PT 1806 guanine etc.
PS Disclosure: fig 1: 6pp; Japanese.
CC This sequence encodes a human sloan kettering institute (SKI) related
CC protein. A truncated protein can also be encoded by bases 709-1806.
CC The protein can be used as a cancer marker by preparing an antibody
CC against it. The copy no. of a cancer gene can be found by Southern
CC blotting using the gene or fragments. The degree of malignancy of a
CC cancer can thus be determined. See also 003743.
SQ Sequence 2839 BP; 936 A; 521 C; 586 G; 796 T;

Query Match      80.0%; Score 80; DB 1; Length 2839;
Best Local Similarity 56.5%; Pred. No. 7,71e+01;
Matches 13; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 1420 accttcctcaaatgtagcgt 1442
||:||||:||||:||||
Cc 1 ACNTNCAYCABAAYGCMAYGT 23

RESULT
8
ID 003743 standard; DNA; 2888 BP.
AC 003743;
DT 15-AUG-1989 (first entry)
DE Human SKI related gene
KW Human sloan kettering institute related gene; cancer; ss.
FW Key Location/Qualifiers
FT cds 709..1956
/*tag= a
/label=human SKI related protein

FN J02053485-A.
PD 22-FEB-1990.
PR 19-AUG-1988; 205906.
PR 19-AUG-1988; JP-205906.
PA (RIKA) Rikagaku Kenkyusho .
DR WPI; 90-103117-A.
P-PSDB: R03665.
PT Human SKI related gene - coded by base sequence including 709 adenine to
PT 1806 guanine etc.
PS Disclosure: fig 2; 6pp; Japanese.
CC This sequence encodes a human sloan kettering institute (SKI) related
CC protein. A truncated protein can also be encoded by bases 709-1806.
CC The protein can be used as a cancer marker by preparing an antibody
CC against it. The copy no. of a cancer gene can be found by Southern
CC blotting using the gene or fragments. The degree of malignancy of a
CC cancer can thus be determined. See also 003743.
SQ Sequence 2839 BP; 936 A; 521 C; 586 G; 796 T;

```


Query Match 75.0% Score 75; DB 13; Length 339;
 Best Local Similarity 63.2% Pred. No. 2.09e+02;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 95 acgttctatctgtagga 113
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYTGRTGNA 5

RESULT 12
 ID 068134 standard; DNA; 339 BP.
 AC 068134;
 DT 26-JUN-1995 (first entry)
 DE Sequence encoding human interleukin-3 muteln (pMon13367).
 KW Human Interleukin-3; hIL-3; mutant; muteln; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 DR P-PSDB; R56108.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia.
 PS Claim 42; Page 150; 388pp; English.
 CC The sequence encodes a muteln of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The mutelns may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The mutelns are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the mutelns are given in GENSEQ Accession Numbers
 CC 068083 - 068189.

SQ Sequence 339 BP; 99 A; 91 C; 69 G; 80 T;

Query Match 75.0% Score 75; DB 13; Length 339;
 Best Local Similarity 63.2% Pred. No. 2.09e+02;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

DB 95 acgttctatctgtagga 113
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYTGRTGNA 5

RESULT 13
 ID 068135 standard; DNA; 339 BP.
 AC 068135;
 DT 26-JUN-1995 (first entry)
 DE Sequence encoding human interleukin-3 muteln (pMon13369).
 KW Human Interleukin-3; hIL-3; mutant; muteln; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.

DR P-PSDB; R56109.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 42; Page 150-151; 388pp; English.
 CC The sequence encodes a muteln of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The mutelns may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The mutelns are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the mutelns are given in GENSEQ Accession Numbers
 CC 068083 - 068189.

SQ Sequence 339 BP; 102 A; 93 C; 66 G; 78 T;

Query Match 75.0% Score 75; DB 13; Length 339;
 Best Local Similarity 54.5% Pred. No. 2.09e+02;
 Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

DB 95 acatgacatctgtaggacg 116
 ||:| | :||:| | | |
 CP 23 ACRTTNCRTTYTGRTGNANG 2

RESULT 14
 ID 068115 standard; DNA; 339 BP.
 AC 068115;
 DT 08-AUG-1995 (first entry)
 DE Sequence encoding human interleukin-3 muteln (pMon13300).
 KW Human Interleukin-3; hIL-3; mutant; muteln; ss.
 OS Synthetic.
 PN W09412638-A.
 PD 09-JUN-1994.
 PF 22-NOV-1993; U11197.
 PR 24-NOV-1992; US-981044.
 PA (MONS) MONSANTO CO.
 PI (SEAR) SEARLE & CO G D.
 PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
 PI Easton AM, Klein BK, McKearn JP, Ollins PO, Paik K;
 PI Thomas JW;
 DR WPI; 94-200266/24.
 PT New human interleukin-3 mutant polypeptide(s) - useful for
 PT treating various blood cell deficiencies, including leukopenia,
 PT neutropenia and leukaemia
 PS Claim 42; Page 145-146; 388pp; English.
 CC The sequence encodes a muteln of hIL-3, comprising Met-Ala-
 CC (15-125)hIL-3. The mutelns may have biological activities similar
 CC to or better than hIL-3 and, in some cases, may also have an
 CC improved side effect profile, e.g. better therapeutic index. They
 CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
 CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
 CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
 CC haematopoietic deficiency associated with treatment with AZT or due
 CC to dialysis; or immunodeficiency resulting from viral infection,
 CC exposure to radiation or cancer treatment.
 CC The mutelns are made by recombinant DNA techniques. Recombinant DNA
 CC sequences encoding the mutelns are given in GENSEQ Accession Numbers
 CC 068083 - 068189.

```

RESULT 15
ID 068146 standard; DNA; 339 BP.
AC 068146;
DT 26-JUN-1995 (first entry)
DE Sequence encoding human interleukin-3 mutein (pMon13383).
KW Human interleukin-3; hIL-3; mutant; mutein; ss.
OS Synthetic.
PN MO9412638-A.
PD 09-JUN-1994.
PF 22-NOV-1993; U11197.
PR 24-NOV-1992; US-981044.
PA (MONS ) MONSANTO CO.
PA (SEAR ) SEARLE & CO G D.
PI Abrams MA, Bauer SC, Braford-Goldberg SR, Caparon MH;
PI Easton AM, Klein BK, McKearn JP, Olin PO, Paik K;
PI Thomas JW;
DR WPI; 94-200266/24.
DR P-PSDB; R56120.
PT New human interleukin-3 mutant polypeptide(s) - useful for
PT treating various blood cell deficiencies, including leukopenia,
PT neutropenia and leukaemia
PS Claim 42; Page 153; 388pp; English.
CC The sequence encodes a mutein of hIL-3, comprising Met-Ala-
CC (15-125)hIL-3. The muteins may have biological activities similar
CC to or better than hIL-3 and, in some cases, may also have an
CC improved side effect profile, e.g. better therapeutic index. They
CC may be used to treat e.g. leukopenia, neutropenia, thrombocytopenia,
CC (aplastic) anemia, Chediak-Higashi syndrome, SLE, leukaemia,
CC myelodysplastic syndrome, myelofibrosis; bone marrow suppression or
CC haematopoietic deficiency associated with treatment with AZT or due
CC to dialysis; or immunodeficiency resulting from viral infection,
CC exposure to radiation or cancer treatment.
CC The muteins are made by recombinant DNA techniques. Recombinant DNA
CC sequences encoding the muteins are given in GENESEQ Accession Numbers
CC 068083 - 068189.
CC Sequence 339 BP; 101 A; 92 C; 69 G; 77 T;

Query Match 75.0%; Score 75; DB 13; Length 339;
Best Local Similarity 63.2%; Pred. No. 2.09e+02;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 95 acgttcattctgataaga 113
   ||:|:| | :||:|:|:| |
CP 23 ACRTTNCRCRTYTGRTGNA 5

```

Search completed: Tue Apr 27 10:35:39 1999
 Job time : 28 secs.

(TM)

itchm

1000

6

mb15

enba

7

-115

ATF13C5

| trans | |

```

comes from Fig. 4"
/codon_start=1
/product="ORF1 receptor agonist precursor"
/db_xref="PID:91172239"
/translation="LSSVSSCPEDLTCQERLHPAPGSFNKILICCEKVPRL
WLTCTKAMASDSLESPADPELTSAIYOSKASEMHRKRSVSYOARAEAD
AEEVDEADEVDEKOLKRFEGFTGARAKRLANOKRSEFMROIYLVSMOSSORR
TLHONGNV"
BASE COUNT      222 a      271 c      237 g      195 t
ORIGIN
Query Match      99.0%; Score 923; DB 29; Length 925;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 924; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      1 CTCCTCCAGCGTGTTCAGACCTGTCGCCGAGAGATGCTCTCACTGCCAGAGAGGCTCCAC 60
QY      1 CTCCTCCAGCGTGTTCAGACCTGTCGCCGAGAGATGCTCTCACTGCCAGAGAGGCTCCAC 60
QY      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
Db      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
QY      61 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
Db      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
QY      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
Db      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
QY      121 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATCTCCATGTGAAGAGAGGCTTTC 120
Db      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
QY      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
Db      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
QY      181 CCTGCTGATCCAGAGCTCAGCTCCGCTCTTACAGTGGAAAGCCTCGGAGATGAC 240
Db      241 CACCTGAAGAGATGCGCGCTGTCAAGAGTGTGTGCAAGCCGAGAGAGAGAGAGGCTGAG 300
QY      241 CACCTGAAGAGATGCGCGCTGTCAAGAGTGTGTGCAAGCCGAGAGAGAGAGAGGCTGAG 300
Db      301 GCAGATGAGAGAGCTGTGCGAGATGAGCGCATGAGTGGAGAGAGAGAGAGAGAGAGAGAG 360
QY      301 GCAGATGAGAGAGCTGTGCGAGATGAGCGCATGAGTGGAGAGAGAGAGAGAGAGAGAGAG 360
Db      361 AGGTTGGGGGCTTCACTGGGGGCGGAGAGTCAAGCCGAGAGAGAGAGAGAGAGAGAGAG 420
QY      361 AGGTTGGGGGCTTCACTGGGGGCGGAGAGTCAAGCCGAGAGAGAGAGAGAGAGAGAGAG 420
Db      421 TTCAGTGAATTTATGAGGAGTACCTGCTGTGAGCATGTGCAAGCCGAGAGAGAGAGAG 480
QY      421 TTCAGTGAATTTATGAGGAGTACCTGCTGTGAGCATGTGCAAGCCGAGAGAGAGAGAG 480
Db      481 ACTCTGACACAGAAATGTAATGTGTAGCCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 540
QY      481 ACTCTGACACAGAAATGTAATGTGTAGCCAGAAAGAGAGAGAGAGAGAGAGAGAGAG 540
Db      541 TGCAGCCATGAGATGAGTACAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY      541 TGCAGCCATGAGATGAGTACAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db      601 GCGCGGAGAGTCAAGATGCTCTTCCCTGAGGAGTGAACACCCGCGGAGAGAGAGAG 660
QY      601 GCGCGGAGAGTCAAGATGCTCTTCCCTGAGGAGTGAACACCCGCGGAGAGAGAGAG 660
Db      661 GCATGCTCCACCAATCTCTGCTCAATCTCTGCTCAATCTCTGCTCAATCTCTGCTCAAT 720
QY      661 GCATGCTCCACCAATCTCTGCTCAATCTCTGCTCAATCTCTGCTCAATCTCTGCTCAAT 720
Db      721 ACATTTTAAATGAGCCCATCTTCTGCTCAATCTCTGCTCAATCTCTGCTCAATCTCTG 780
QY      721 ACATTTTAAATGAGCCCATCTTCTGCTCAATCTCTGCTCAATCTCTGCTCAATCTCTG 780
Db      781 GAACATGAATATGAGCTGGGTTTGGCTCAACCACTGCCAATATGTTTGAATGTAAGAG 840
QY      781 GAACATGAATATGAGCTGGGTTTGGCTCAACCACTGCCAATATGTTTGAATGTAAGAG 840
Db      841 CTGTTCTTTTACATGATGTTGAAACAATTTTCCATTAACTTACTAGACAAA 900
QY      841 CTGTTCTTTTACATGATGTTGAAACAATTTTCCATTAACTTACTAGACAAA 900

```

```

QY      841 CTGTTCTTTTACATGATGTTGAAACAATTTTCCATTAACTTACTAGACAAA 900
Db      901 TGGTTAATAAAAAAAAAAAAAAAAAA 925
QY      901 TGGTTAATAAAAAAAAAAAAAAAAAA 925

RESULT 2
LOCUS      RNPmRNA      988 bp      RNA      ROD      10-SEP-1996
DEFINITION R.norvegicus mRNA for prepronociceptin.
ACCESSION      X97375
NID      91532064
KEYWORDS      prepronociceptin.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE      1 (bases 1 to 988)
AUTHORS      Mollereau,C., Simons,M.J., Soularue,P., Lhiers,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE      Structure, tissue distribution, and chromosomal localization of the
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
MEDLINE      96323281
REFERENCE      2 (bases 1 to 988)
AUTHORS      Parmentier,M.
TITLE      Direct Submission
JOURNAL      Submitted (19-APR-1996) M. Parmentier, Universite Libre de
Bruxelles, 1 R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
COMMENT      Related sequence 579730.
FEATURES      Location/Qualifiers
source      1..988
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="lambda DASHII (Stratagene #943501)"
/clone_1b="lambda ZAPRI (brain cDNA)"
/clone="RGE109"
/clone="RB101"
<1..150
/number=2
join<1..150,151..615,616..974)
25..69
join(25..150,151..570)
CDS
mRNA
sig_peptide
exon
<1..150
/number=2
join<1..150,151..615,616..974)
25..69
join(25..150,151..570)
CDS
mat_peptide
exon
151..615
/number=3
616..974
/number=4
971
BASE COUNT      231 a      292 c      250 g      215 t
ORIGIN
Query Match      99.0%; Score 923; DB 29; Length 988;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 924; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      64 CTCCTCAGAGTGTTCAGAGGTCGCCGAGAGATGCTCTCACTGCCAGAGAGGCTCCAC 123
QY      1 CTCCTCAGAGTGTTCAGAGGTCGCCGAGAGATGCTCTCACTGCCAGAGAGGCTCCAC 60
Db      124 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATGTGAAGAGAGGCTTTC 183
QY      124 CCGGCTCCGGGAGCTTCAACCTGAAGCTGTGCATCTCCATGTGAAGAGAGGCTTTC 183

```

```

OY      61  CCGGCTCCGGGACGCTTCAACCTGAACTGTCATCTCTCAAGTGTGAAGAGAGCTTTC 120
Db      184  CCGGGCCCTCTCTGAGCTCTTTGCAACCAAGCCATGGCCAGTGAAGTCTGAGAGCTCAGC 243
OY      121  CCGGGCCCTCTCTGAGCTCTTTGCAACCAAGCCATGGCCAGTGAAGTCTGAGAGCTCAGC 180
Db      244  CCTGCTGATCCAGAGCTACGCTGCTCTCTTTACAGTGGAAAGCTCGAGATGAG 303
OY      181  CCTGCTGATCCAGAGCTACGCTGCTCTCTTTACAGTGGAAAGCTCGAGATGAG 240
Db      304  CACCTGAAGAGATGCCCGCTGTGAGAGTGTGTGCAAGCCGAGAGCAGAGCCTGAG 363
OY      241  CACCTGAAGAGATGCCCGCTGTGAGAGTGTGTGCAAGCCGAGAGCAGAGCCTGAG 300
Db      364  GCAGATGAGAGAGCTGTGAGAGTGGCCGATGAGTGGAGAGAGAGAGAGAGAGAG 423
OY      301  GCAGATGAGAGAGCTGTGAGAGTGGCCGATGAGTGGAGAGAGAGAGAGAGAGAG 360
Db      424  AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAGCCGGAAGTTGGCCAGCAGAGAGG 483
OY      361  AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAGCCGGAAGTTGGCCAGCAGAGAGG 420
Db      484  TTCAGTGAATTTATGAGCAGTACCTGCTCTGAGCAGTCAAGTCAAGCCAGCCGCGC 543
OY      421  TTCAGTGAATTTATGAGCAGTACCTGCTCTGAGCAGTCAAGTCAAGCCAGCCGCGC 480
Db      544  ACTTGCACCAAGATGTAATGTGTAGCCAGAGAGAGCCCTCCAGCTGCACGGCCAC 603
OY      481  ACTTGCACCAAGATGTAATGTGTAGCCAGAGAGAGCCCTCCAGCTGCACGGCCAC 540
Db      604  TGCACCCATGAGATGAGTGAAGTGAAGCCCGGAGACAGATGTCACCAACCAAGACTG 663
OY      541  TGCACCCATGAGATGAGTGAAGTGAAGCCCGGAGACAGATGTCACCAACCAAGACTG 600
Db      664  GCGCGGAGTCAAGATCTCTCTCTCTGAGGAGTGAAGACCCGCGAGCTCCCGACA 723
OY      601  GCGCGGAGTCAAGATCTCTCTCTCTGAGGAGTGAAGACCCGCGAGCTCCCGACA 660
Db      724  GCATGTCTCACCACATCTCTGTGTATCATCAGAGTGTATTTTGTAAATCTCCAGCTA 783
OY      661  GCATGTCTCACCACATCTCTGTGTATCATCAGAGTGTATTTTGTAAATCTCCAGCTA 720
Db      784  ACATTTAATGGCCCATCTTCTGTCATCTGTCCTGCGCTCGTGAAGGCGAGTGAAG 843
OY      721  ACATTTAATGGCCCATCTTCTGTCATCTGTCCTGCGCTCGTGAAGGCGAGTGAAG 780
Db      844  GAACATGAATCAGACCTGCGGTTTGGCTCAGCAGTGCATACCTGCTGTTGTAAGAG 903
OY      781  GAACATGAATCAGACCTGCGGTTTGGCTCAGCAGTGCATACCTGCTGTTGTAAGAG 840
Db      904  CTGTTCTTTTGTGATGATGTTTGAAGAACTTCTCATTAAGTCTACTGAGCAAAA 963
OY      841  CTGTTCTTTTGTGATGATGTTTGAAGAACTTCTCATTAAGTCTACTGAGCAAAA 900
Db      964  TGGTTAATAAAAAAGAAAAAAGAAAA 988
OY      901  TGGTTAATAAAAAAGAAAAAAGAAAA 925

RESULT 3  RNU48262  972 bp  mRNA  ROD  23-AUG-1996
LOCUS    Ratius norvegicus pre-pro-orphanin FQ mRNA, complete cds.
DEFINITION
ACCESSION U48262
NID 91185011
KEYWORDS
SOURCE Norway rat.
ORGANISM Ratius norvegicus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Rattus.
REFERENCE 1 (bases 1 to 972)
AUTHORS Nothacker, H.-P., Reinscheid, R.K., Mansour, A., Henningsen, R.A.,
Artacl, A., Monsma, F.J. Jr., Watson, S.J. and Clivelli, O.

```

```

TITLE Primary structure and tissue distribution of the orphanin FQ
JOURNAL Precursor
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
MEDLINE 96323283
REFERENCE 2 (bases 1 to 972)
AUTHORS Nothacker, H.-P. and Henningsen, R.A.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202,
Hofmann-La Roche AG, Grenzachstr 124, Basel, 4070, Switzerland
FEATURES
source
1..972
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
slg_peptide
28..84
CDS
85..570
/codon_start=1
/product="pre-pro-orphanin FQ"
/db_xref="pid:g1185012"
/translation="MKILFDVLLSLSSVFSSEPCEDLTCQERLHPAGSFYKLC
ILCOEIVFPRPLTCTKMAASDSBLSFADPELTSALYQSASMOHLKRPVR
SVQARDAPEADAEVADAEVOKQLOKRGEGTGARKSARKLANORFSEFMQ
YLVLSMQSSQRRRLHQNGNV"
mat_peptide
85..570
/product="pro-orphanin FQ"
3'UTR
574..972
polya_site
972
BASE COUNT 214 a 293 c 251 g 214 t
ORIGIN
Query Match 97.0%; Score 904; DB 29; Length 972;
Best Local Similarity 99.9%; Pred. No. 0.00e+00;
Matches 905; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 67 CTCCTCAGCGTGTTCAGAGCTGTCCGAGAGTACCTGCTCAGCTGCGAGAGAGCTCCAC 126
OY 1 CTCCTCAGCGTGTTCAGAGCTGTCCGAGAGTACCTGCTCAGCTGCGAGAGAGCTCCAC 60
Db 127 CCGGCTCCGGGACAGCTTCAACCTGAAGTGTGATCTCTCAGTGTGAAGAGGCTCTTC 186
OY 61 CCGGCTCCGGGACAGCTTCAACCTGAAGTGTGATCTCTCAGTGTGAAGAGGCTCTTC 120
Db 187 CCGGCTCCCTCTCTGAGCTCTTTGCAACCAAGCCATGGCCAGTGAAGTGAAGCTCAGC 246
OY 121 CCGGCTCCCTCTCTGAGCTCTTTGCAACCAAGCCATGGCCAGTGAAGTGAAGCTCAGC 180
Db 247 CCGTGTATCCAGAGCTGAGTCCGCTGCTCTTACAGAGTGAAGAGCTCGAGATGAG 306
OY 181 CCGTGTATCCAGAGCTGAGTCCGCTGCTCTTACAGAGTGAAGAGCTCGAGATGAG 240
Db 307 CACCTGAAGAGATGCCCGCTGTGAGAGTGTGTGAAGCCGAGAGCAGAGAGCTGAG 366
OY 241 CACCTGAAGAGATGCCCGCTGTGAGAGTGTGTGAAGCCGAGAGCAGAGAGCTGAG 300
Db 367 GCAGATGAGAGAGCTGTGAGAGTGGCCGATGAGTGAAGCAGAGAGCAGCTGAGAAA 426
OY 301 GCAGATGAGAGAGCTGTGAGAGTGGCCGATGAGTGAAGCAGAGAGCAGCTGAGAAA 360
Db 427 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAGCCGGAAGTTGGCCAGCAGAGAGG 486
OY 361 AGGTTGGGGGCTTCACTGGGGCCGGAAGTCAGCCGGAAGTTGGCCAGCAGAGAGG 420
Db 487 TTCAGTGAATTTATGAGCAGTACCTGCTCTGAGCAGTCAAGTCAAGCCAGCCGCGC 546
OY 421 TTCAGTGAATTTATGAGCAGTACCTGCTCTGAGCAGTCAAGTCAAGCCAGCCGCGC 480
Db 547 ACTTGCACCAAGATGTAATGTGTAGCCAGAGAGAGCCCTCCAGCTGCACGGCCAC 606
OY 481 ACTTGCACCAAGATGTAATGTGTAGCCAGAGAGAGCCCTCCAGCTGCACGGCCAC 540
Db 607 TGCACCCATGAGATGAGTGAAGTGAAGCCCGGAGACAGATGTCACCAACCAAGACTGCA 666

```

QY 541 TGCAACCATGAGCATCCAGGTGAGCCGCCCTACAGCATGTGTCCACACCAAGACTGCA 600
 Db 667 GCGGGAGTACAGATTCCTCTCCCTGAGGACAGTACACCCGCGGACCTCCACCA 726
 QY 601 GGGGGGAGTAGATTCTCTCTCCCTGAGGACAGTACACCCGCGGACCTCCACCA 660
 Db 727 GCATGTCTACACCAATCTGTGTCTCATCATGAGTATTTTGTATTTCTCCAGCTA 786
 QY 661 GCATGTCTACACCAATCTGTGTCTCATCATGAGTATTTTGTATTTCTCCAGCTA 720
 Db 787 ACATTTTAATGCCCCATCTCTGTCTCATCTGCCCCCTGTGAGGCGGAGTGAAG 846
 QY 721 ACATTTTAATGCCCCATCTCTGTCTCATCTGCCCCCTGTGAGGCGGAGTGAAG 780
 Db 847 GAACATGAATTCAGACTGGGGTTTGGCCCTACACCATCTGCATTAAGTGTGAAGAG 906
 QY 781 GAACATGAATTCAGACTGGGGTTTGGCCCTACACCATCTGCATTAAGTGTGAAGAG 840
 Db 907 CTGTCTTTTGAAGTATTTGAAACAACATTTCTCATTAACCTTCTACTGAGCAAA 966
 QY 841 CTGTCTTTTGAAGTATTTGAAACAACATTTCTCATTAACCTTCTACTGAGCAAA 900
 Db 967 TGGTTA 972
 QY 901 TGGTTA 906

RESULT 4
 LOCUS MUSNOP 1354 bp mRNA
 DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
 ACCESSION D82866
 NID 91311472
 KEYWORDS nociceptin/orphanin FQ.
 SOURCE Mus musculus adult brain cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1354)
 Takeshima, H.
 Direct Submission
 Submitted (25-DEC-1995) to the DDBJ/EMBL/GenBank databases. Hiroshi
 Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of
 Medicine, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
 (E-mail: takeshimem.u-tokyo.ac.jp, Tel: 03-3812-2111 (ex. 3422),
 Fax: 03-3815-9360)
 2 (bases 1 to 1354)
 Takeshima, H.
 Unpublished (1996)
 3 (sites)
 Houtani, T., Nishi, M., Takeshima, H., Nukada, T. and Sugimoto, T.
 Structure and regional distribution of nociceptin/orphanin FQ
 precursor
 Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)
 96216718
 MEDLINE
 JOURNAL
 FEATURES
 source
 Location/Qualifiers
 1..1354
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /dev_stage="adult"
 /tissue_type="brain"
 360..923
 /codon_start=1
 /product="nociceptin/orphanin FQ"
 /db_xref="PID:d1012281"
 /translation="MKILFCDVLLSLSSVSSPRLCTOELHPAPDSFNKTC
 ILQCEKVPRLMTVCTKVMASGSGQLSPADPEVSAALVTPKASEQHLKMPVR
 STVOVDABEGADAEPADEPADEAEVEEQKQKQKRGFTGARKSARKLANQKRF
 SEPDKYLVLSQSSQSRRTLHQNMY"
 1317..1322

PolyA_signal
 BASE COUNT 309 a 370 c 386 g 289 t
 ORIGIN

Query Match 59.4%; Score 554; DB 29; Length 1354;
 Best Local Similarity 89.7%; Fred. No. 0.00e+00;
 Matches 858; Conservative 0; Mismatches 58; Indels 41; Gaps 14;

Db 399 CTCTCCAGCGTTTCAGAGCTGTCCAGAGGATGCTCCACCTGCGGAGGAGAGTCCAC 458
 QY 1 CTCTCCAGCGTTTCAGAGCTGTCCAGAGGATGCTCCACCTGCGGAGGAGTCCAC 60
 Db 459 CCAGTCCAGACACTTCACTTAAGAGCTGCAATCCATGATGTAAGAGAGTCTTC 518
 QY 61 CCGGCTCGGGCACTTCAACTCAACTGATGATCTCCAGTGAAGAGAGTCTTC 120
 Db 519 CCGGCGCTCTCTGAGCTGATGACCAAGTATGAGCCAGTGGCTCCGCGGAGTCCAG 578
 QY 121 CCGGCGCTCTCTGAGCTGATGACCAAGTATGAGCCAGTGGCTCCGCGGAGTCCAG 180
 Db 579 CTGTCTGACCCAGAGCTTGTGAGCTGCTTTTACAGCCAAAGGCTCGGAGATGAG 638
 QY 181 CTGTCTGATCCAGAGCTCAAGTCCGCTGCTTTACAGTGCAGAACTTCGGAGATCAG 240
 Db 639 CACCTGAAGAGAAAGCCGCTGTCGAGCTTGTGCAAGTGTGCAAGTGCAGAGCTGCG 698
 QY 241 CACCTGAAGAGAAAGCCGCTGTCAGAGTGTGTCAGAGCCGAGAGCAGAGCTGAG 300
 Db 699 GCAGATGCTGAGCTGAGCGAGATGCTGAGAGCTGCGGAGATGAGAGAGTGGAG 758
 QY 301 GCAGATGAGAGAGTGTGCGAGATG---AG---GC---GA---TGAG---TGAG 342
 Db 759 CAGAGAGCTGCGAGAAAGGTTGGGGCTTCACCGGGGCGGAGAAATCACCGCGAG 818
 QY 343 CAGAGAGCTGCGAGAAAGGTTGGGGCTTCACCGGGGCGGAGAAATCACCGCGAG 402
 Db 819 TTGGCCAAACAGAGCGGTTGAGTATGAGGAGTACTGCTGAGCATGAG 878
 QY 403 TTGGCCAAACAGAGCGGTTGAGTATGAGGAGTACTGCTGAGCATGAG 462
 Db 879 TCAAGTCAAGCGCGGCGACCTGACCCAGCAAGTGTATGTATGCCAGAGAGACCCCT 938
 QY 463 TCAAGTCAAGCGCGGCGACCTGACCCAGCAAGTGTATGTATGCCAGAGAGACCCCT 522
 Db 939 CCGAGCTGACCGGCGACCTGACCCAGCAAGTGTATGTATGCCAGAGAGATG 998
 QY 523 CCGAGCTGACCGGCGACCTGACCCAGCAAGTGTATGTATGCCAGAGATG 582
 Db 999 TCCACTCCAAGACTGCAAGCGCGGAGTCAAGATTCGACCCCGGAGAGAGAGAC 1058
 QY 583 TCCACTCCAAGACTGCAAGCGCGGAGTCAAGATTCGACCCCGGAGAGAGAGAC 638
 Db 1059 CCACAGACACTGCGCGGAGTGTGCGCTCCCTCATGAGCATGTCACACAGATCT 1118
 QY 639 ---A-CAC---C-GC-----G--GCACCTCC-CACACAGTGTCCACCAATCT 680
 Db 1119 GTTGCTACATCAGAGTATTTTGTATTTCTCCAGTATCAATTTAAGGCCCATCT 1178
 QY 681 GTTGCTACATCAGAGTATTTTGTATTTCTCCAGTATCAATTTAAGGCCCATCT 740
 Db 1179 TCTCACTATCTTTCGCCCTCATAGGCGGAGTGAAGAGATCAATTAAGAGCTG 1238
 QY 741 TCTTCTATCTCTGCGCTCTGTAAGGCGGAGTGAAGAGATCAATTAAGAGCTG 800
 Db 1239 GGTTTGGCTCTCAGTCCATTAAGTGTGTAAGAGAGCTGTTTGTGAGTATG 1298
 QY 801 GGTTTGGCTCTCAGTCCATTAAGTGTGTAAGAGAGCTGTTTGTGAGTATG 860
 Db 1299 TTAAACAATTTCTCATTAAGTGTGTAAGAGAGCTGTTTGTGAGTATG 1354
 QY 861 TTGGAACAATTTCTCATTAAGTGTGTAAGAGAGCTGTTTGTGAGTATG 917

RESULT 5
 LOCUS MUSN23K 1047 bp mRNA
 DEFINITION Mouse N23K mRNA for N23K protein, complete cds. 14-MAR-1997


```

/clone.lib="lambda FIXII (Stratagene #946305)"
/clone="MG43"
<1..150
/number=2
join(<1..150,151..>633)
join(25..150,151..588)
/codon_start=1
/product="prepronocleptin"
/db_xref="PID:e24874"
/translation="MKILFCDVLLSLISVSSPCPDCLTQEKIHPADSPNLKTC
ILQCEKVEFPPRLWTCTKYVAGSGGSLSPADPELVSAALYKQKASQMLKMPYR
SLVOYRDAEPGADAPGADAEVQKQLOKRRFGFTGARKSARKLANQKRF
SEFMQYLVLSMSSQRRRLHONGNY"
sig_peptide
mat_peptide
25..69
70..585
/product="prepronocleptin"
/number=3
151..>633
exon
BASE COUNT      138 a      194 c      182 g      119 t
ORIGIN
Query Match      39.3%; Score 366; DB 29; Length 633;
Best Local Similarity 90.0%; Pred. No. 5,37e-201;
Matches 513; Conservative 0; Mismatches 39; Indels 18; Gaps 6;

Db 64 CTCTCAGCGGTTCAGAGCTGTCCAGGAGCTCCTCAGCTGAGGAGAAAGCTCCAC 123
    |||
QY 1 CTCTCAGCGGTTCAGAGCTGTCCAGGAGCTCCTCAGCTGAGGAGAAAGCTCCAC 60
    |||
Db 124 CCAGCTCCAGAGCTTCAACTTAAGAGCTGCATCTCCAGTGTGAAGAGGTCTTC 183
    |||
QY 61 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 120
    |||
Db 184 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 243
    |||
QY 121 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 180
    |||
Db 244 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 303
    |||
QY 181 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 240
    |||
Db 304 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 363
    |||
QY 241 CCGGCTCCGGGACAGCTCAACCTGAAGCTGCATCTCCAGTGTGAAGAGGTCTTC 300
    |||
Db 364 GCAGATGCTGAGCGCTGCGAGATGCTGCGAGATGCTGCGAGATGCTGCGAGATG 423
    |||
QY 301 GCAGATGCTGAGCGCTGCGAGATGCTGCGAGATGCTGCGAGATGCTGCGAGATG 342
    |||
Db 424 CAGAGCAGCTGCAGAAAGGTTTGGGGCTTACCGGGGCGGAAATYACAGCCCGGAG 483
    |||
QY 343 CAGAGCAGCTGCAGAAAGGTTTGGGGCTTACCGGGGCGGAAATYACAGCCCGGAG 402
    |||
Db 484 TTGGCCAAACAGAGGTTTCAAGTATTAAGAGCAAGTACCTGCTGAGCATGAG 543
    |||
QY 403 TTGGCCAAACAGAGGTTTCAAGTATTAAGAGCAAGTACCTGCTGAGCATGAG 462
    |||
Db 544 TCAAGTCAAGCGCGGCGAGCTGACAGCAAGTATTAAGTATTAAGTATTAAGTATTA 603
    |||
QY 463 TCAAGTCAAGCGCGGCGAGCTGACAGCAAGTATTAAGTATTAAGTATTAAGTATTA 522
    |||
Db 604 CCGAGCTGCAGCGGCGAGCTGACAGCAAGTATTAAGTATTAAGTATTAAGTATTA 633
    |||
QY 523 CCGAGCTGCAGCGGCGAGCTGACAGCAAGTATTAAGTATTAAGTATTAAGTATTA 552
    |||
RESULT
LOCUS      7      MUSN27K      972 bp      mRNA
DEFINITION Mouse N23K mRNA for developmental-regulated molecule, complete cds.
ACCESSION D50055
NID        91262587
KEYWORDS   developmental-regulated molecule; N23K; neuropeptide.

```

```

SOURCE
ORGANISM    Mus musculus neuroblastoma cell_line: NS20Y cDNA to mRNA.
REFERENCE
AUTHORS     Saito, Y., Maruyama, K., Kawano, H., Hagiino-Yamagishi, K., Kawamura, K.,
            Saito, Y., and Kawashima, S.
TITLE       Molecular cloning and characterization of a novel form of
            neuropeptide gene as a developmentally regulated molecule
            J. Biol. Chem. 271 (26), 15615-15622 (1996)
REFERENCE
AUTHORS     Saito, Y.
TITLE       Unpublished (1996)
REFERENCE
AUTHORS     Saito, Y.
TITLE       Submitted (05-APR-1995) to the DDBJ/EMBL/GenBank databases. Yumiko
            Saito, Tokyo Metropolitan Institute for Medical Science, molecular
            biology, Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan
            (E-mail: saito@metinstoken.or.jp, Tel: 03-3823-2101 (ex. 5248),
            Fax: 03-5685-6609)
FEATURES
            location/Qualifiers
            1..972
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /cell_line="NS20Y"
            /cell_type="neuroblastoma"
            /gene="N23K"
            /gene="N23K"
            /gene="N23K"
            /codon_start=1
            /product="developmental-regulated molecule"
            /db_xref="PID:d1009394"
            /translation="MKILFCDVLLSLISVSSPCPDCLTQEKIHPADSPNLKTC
            ILQCEKVEFPPRLWTCTKYVAGSGGSLSPADPELVSAALYKQKASQMLKMPYR
            SLVOYRDAEPGADAPGADAEVQKQLOKRRFGFTGARKSARKLANQKRF
            SEFMQYLVLSMSSQRRRLHONGNYIYIPRTACVHSKTRPVRIPSPRH"
            972
BASE COUNT      213 a      291 c      246 g      222 t
ORIGIN
Query Match      38.1%; Score 355; DB 29; Length 972;
Best Local Similarity 83.5%; Pred. No. 4,23e-194;
Matches 773; Conservative 0; Mismatches 109; Indels 53; Gaps 27;

Db 82 CTCTCAGCGGTTCAGAGCTGTCCAGGAGCTCCTCAGCTGAGGAGAAAGCTCCAC 141
    |||
QY 1 CTCTCAGCGGTTCAGAGCTGTCCAGGAGCTCCTCAGCTGAGGAGAAAGCTCCAC 60
    |||
Db 142 CCAGCTCCAGAGCTTCAACTTAAGAGCTGCATCTCCAGTGTGAAGAGGTCTTC 201
    |||
QY 61 CCGGCTCCGGGACAGCTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGGTCTTC 120
    |||
Db 202 CCGGCTCCGGGACAGCTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGGTCTTC 261
    |||
QY 121 CCGGCTCCGGGACAGCTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGGTCTTC 180
    |||
Db 262 CCGGCTCCGGGACAGCTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGGTCTTC 321
    |||
QY 181 CCGGCTCCGGGACAGCTCAACCTGAAGCTGTGCATCTCCAGTGTGAAGAGGTCTTC 240
    |||
Db 322 CAGCTGAAGAGAAAGCGGCTGCGAGCTTGTGCAAGTGCAGATGCAAGAGCTTGC 381
    |||
QY 241 CAGCTGAAGAGAAAGCGGCTGCGAGCTTGTGCAAGTGCAGATGCAAGAGCTTGC 300
    |||
Db 382 GCAGATGCTGAGCGCTGCGAGATGCTGAGCGCGAGATGCTGAGCGCGAGATGCTGAG 441
    |||
QY 301 GCAGATGCTGAGCGCTGCGAGATGCTGAGCGCGAGATGCTGAGCGCGAGATGCTGAG 342
    |||

```

Db 442 CAGAGGAGCTGCGAGAAAGGTTGGGGCTTACCGGGGCCGGAATCAGCCCGAG 501
|||||
Oy 343 CAGAGGAGCTGCGAGAAAGGTTGGGGCTTACCGGGGCCGGAATCAGCCCGAG 402
|||||
Ub 502 TTGGCCACAGAGCCGTTGAGTATGAGAGCAGTACTGTTGCTGAGCAGTGCAG 561
|||||
Oy 403 TTGGCCACAGAGCCGTTGAGTATGAGAGCAGTACTGTTGCTGAGCAGTGCAG 462
|||||
Db 562 TCAAGTCAAGCCGGCGGCGCTGACAGAGTGGCA----T--CGAGGTATCCCCG 614
|||||
Oy 463 TCAAGTCAAGCCGGCGGCGCTGACAGAGTGGTATGAGAGCAGAGAGAGCCCT 522
|||||
Db 615 TACAGCATGT--GTCCATCTCAAGACCTG--CAGGCCG--GAGTCAGATTC--C--TCCC 666
|||||
Oy 523 CCCAGCTGACCCGGCGGCGCTGACAGATTCAGAGTGGAGCCCGCTACAGATGTG 582
|||||
Db 667 TCC-C-CGAGGAGCTG--AGCACC-A--CAGCA--CCTCCCTGCCG--GT--TGTG-C 713
|||||
Oy 583 TCCACACCAAGACTGCGAGCCGGGAGTCAAGATTCCTCTCCCTGAGCAGTGAAC 642
|||||
Db 714 CC-CT-C-CCTC---ATGGCATGTCTCACACAGATCCTGTGTACATCAGAGTATTT 767
|||||
Oy 643 CCGGGGACCTCCCGACAGTGTCTCACACAACTGTGTGTACATCAGAGTATTT 702
|||||
Db 768 TTGTATTTCTCCAGCTTAACATTTATGCCCCATCTTCACATCTTCTGCCCTCT 827
|||||
Oy 703 TTGTATTTCTCCAGCTTAACATTTATGCCCCATCTTCACATCTTCTGCCCTCT 762
|||||
Db 828 CATAGGCGCAGTGAAGAGATGAATCAGACCTGGGTTTGGCTTCACAGCCAT 887
|||||
Oy 763 CGTAGGCGCAGTGAAGAGATGAATCAGACCTGGGTTTGGCTTCACAGCCAT 822
|||||
Db 888 AACTGTTTGAAGAGCTGTTTGTGATGATCTGTAAACAATTCTTCATTA 947
|||||
Oy 823 AACTGTTTGAAGAGCTGTTTGTGATGATCTGTAAACAATTCTTCATTA 882
|||||
Db 948 AACTTCTACTAGCAA--TGCTAAT 972
|||||
Oy 883 AACTTCTACTAGCAAATGCTAAT 908
|||||
RESULT 8
LOCUS HSPHNRNA 1015 bp RNA PRI 10-SEP-1996
DEFINITION H.sapiens mRNA for prepronocleptin.
ACCESSION X97370
NID g1531988
KEYWORDS prepronocleptin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euteheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1015)
Molleran,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Muller,J.C. and Parmentier,M.
Structure, tissue distribution, and chromosomal localization of the
prepronocleptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
JOURNAL MEDLINE
REFERENCE 96323281
TITLE 2 (bases 1 to 1015)
AUTHORS Parmentier,M.
JOURNAL Direct Submission
Submitted (19-APR-1996) M. Parmentier, Universite Libre de
Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
FEATURES
SOURCE location/Qualifiers
1..1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="lambda DASHII (Stratagene #94520)"
/clone="HG91"
/clone="HGEX107"
/chromosome="8"
/map="p21"

/map="1.9 cr from WI-1172 marker"
<1..149
/number=2
mRNA join(<1..149,150..601,602..>1015)
sig_peptide 24..68
CDS join(24..149,150..556)
/codon_start=1
/product="prepronocleptin"
/db_xref="PID:e24477"
/db_xref="PID:g1531988"
/translation="MKVLCIDLLLSLFSVSSCORDLTCOEKTLHPALDSFLEVC
ILCEKRPSPPLPTPCRYMASSWOLSPAPRYAALYORASEMHLRPMYR
SLRQEEPEPEPEHEEEMEQOLQKRFGEFTARKSARKLNQKRFSEPMROYLVLS
MOSSQRRRLHONGV"
mat_peptide 69..551
/product="prepronocleptin"
exon 150..601
exon 602..>1015
poly_site 1015
BASE COUNT 212 a 304 c 264 g 235 t
ORIGIN
Query Match 36.1%; Score 336; DB 26; length 1015;
Best Local Similarity 82.5%; Pred. No. 3.33e-182;
Matches 534; Conservative 0; Mismatches 96; Indels 17; Gaps 9;
Db 64 TCTCCAGTGTGTGAGAGAGTGTGACAGAGGAGCTCTCAATCCAGAGAGAGTCCAGC 123
|||||
Oy 2 TCTCCAGTGTGTGAGAGAGTGTGACAGAGGAGCTCTCAATCCAGAGAGAGTCCAGC 61
|||||
Db 124 CAGCCTGACAGAGTGTGACCTGAGAGTGTGACCTCTGAGTGTGAGAGAGTCTTCC 183
|||||
Oy 62 CAGCCTGAGAGAGTGTGACCTGAGAGTGTGACCTCTGAGTGTGAGAGAGTCTTCC 121
|||||
Db 184 CAGCCTGAGAGTGTGACCTGAGAGTGTGACCTCTGAGTGTGAGAGAGTCTTCC 243
|||||
Oy 122 CAGCCTGAGAGTGTGACCTGAGAGTGTGACCTCTGAGTGTGAGAGAGTCTTCC 181
|||||
Db 244 CTGCGGCCCGAGAGAGTGTGAGAGTGTGACCTCTGAGTGTGAGAGAGTCTTCCAGC 303
|||||
Oy 182 CTGCTGATTCAGAGCTGACGCTGCTCTCTTTCAGAGTGAAGAGCTCGAGATGAGC 241
|||||
Db 304 ATCTGCGCGAGATCCCGAGATCCGAGATCTTTCAGAGAGC--AG--GAAGAGCCGAGC 360
|||||
Oy 242 ACCTGAGAGAGATCCCGAGATCTGAGAGTGTGAGAGAGCAGAGAGCTGAGC 301
|||||
Db 361 CTG--GCA----TG--G-AG--GAGGCTGTGATGAGAGAGAGAGAGAGAGAGAG 408
|||||
Oy 302 CAGATGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
|||||
Db 409 GATTTGGGGGCTTACCGGGGCCCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 468
|||||
Oy 362 GATTTGGGGGCTTACCGGGGCCCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 421
|||||
Db 469 TCACTGAGTGTGAG 528
|||||
Oy 422 TCACTGAGTGTGAG 481
|||||
Db 529 CCGTGCACAG 588
|||||
Oy 482 CTCTGACAG 539
|||||
Db 589 CTGAG 648
|||||
Oy 540 CTGAG 599
|||||
Db 649 CGCCTGGAGATCAGAGATCTCTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 695
|||||
Oy 600 AGGCGGGAGAGAGAGATCTCTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 646
|||||


```

RESULT 9
LOCUS HSU48263 1198 bp mRNA PRI 23-AUG-1996
DEFINITION Human pre-pro-orphanin FQ (OFQ) mRNA, complete cds.
ACCESSION U48263
NID 91185009
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1198)
AUTHORS Notthacker,H.P., Reinscheid,R.K., Mansour,A., Henningsen,R.A.,
Ardelt,A., Monsma,F.J., Jr., Watson,S.J. and Civelli,O.
TITLE Primary structure and tissue distribution of the orphanin FQ
precursor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
MEDLINE 96333283
REFERENCE 2 (bases 1 to 1198)
AUTHORS Notthacker,H.-P. and Henningsen,R.A.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1996) Hans-Peter Notthacker, PRPN 69/202,
Hofmann-La Roche AG, Grenzachstr 124, Basel, 4070, Switzerland
FEATURES
source
1..1198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
1..211
/gene="OFQ"
212..742
/gene="OFQ"
212..742
/gene="OFQ"
/codon_start=1
/product="pre-pro-orphanin FQ"
/db_xref="PDB:1185010"
/translotion="MKVLCDLILSLFSSVSSCQRCCLTCOEKLPALDSPLEVC
ILCEEVTFPSPLMTPTCTKVARSSMQLSPAPHPVAAALYOPRASHMOLRRMPYR
SLFOEPEPEGEAGEEMEQKQLQKRFQGTGARKARKARKRKSEMRQYLVLS
MSSORRRTLHONQNV"
212..268
/sig_peptide
/gene="OFQ"
599..649
/mat_peptide
/gene="OFQ"
/product="orphanin FQ"
743..1198
/3'UTR
1198
/polyA_site
/note="29 A nucleotides"
BASE COUNT 244 a 335 c 345 g 273 t 1 others
ORIGIN
Query Match 36.1%; Score 336; DB 28; Length 1198;
Best Local Similarity 82.4%; Pred. No. 3.33e-182;
Matches 533; Conservative 1; Mismatches 96; Indels 17; Gaps 9;
Db 252 TCTCCAGTGTTCAGCAGTTGTGAGAGGACTGTCTACATGCCAGAGAGCTCCACC 311
QY 2 TCTCCAGTGTTCAGCAGTGTCCGAGGACTGTCTACATGCCAGAGAGCTCCACC 61
Db 312 CAGCCCTGGACAGCTTGACCTGAGGTGTGACCTCTCGAGTGGAGAGAGGTCTTC 371
QY 62 CGGCTCCGGGAGCTTCAACTGAAGCTGTGACATCTCCAGGTGAGAGAGGTCTTC 121
Db 372 CCAGCCCTCTGTGACTCATGACCAAGTGTGATGCCAGAGTGTGGAGTCAAGC 431
QY 122 CCGCCCTCTGTGACTCTTTCACCAAGCAGGCAATGACTGTGAGACACTCAGCC 181
Db 432 CTGCGCCCTCAGAGCAGTGTGCTGTCTACACAGCCGAGAGCTTCGAGATCAGC 491
QY 182 CTGCTGTGATCAGAGCTCAGCTCCGCTCTCTTACAGTGAAGAGCTCGAGATCAGC 241
Db 492 ATCTGCGGCAATGCGCGAGTCCGAGCTGTGTTCAGAGAGC-AG--GAAGAGCCGAGC 548
QY 242 ACCTGAAGAGAAATGCGCGCTGTGAGAGTGTGTGCAAGCCCAAGCAGAGAGCTGAGG 301

```

```

Db 549 CTG--GCA-----TC--G-AG--GAGCTGTGATGATGAGCAGAAAGACCTGCAGAGA 596
QY 302 CAGATGCAAGAGCCCTGTGCGAGATGAGCCGATGAGTGTGAGCAGAAAGACCTGCAGAAA 361
Db 597 GATTGGGGGCTTCACCGGGGCGCGGAAGTGGCGAGAGTGGCCATCAACAACCGGT 656
QY 362 GATTGGGGGCTTCACCGGGGCGCGGAAGTGGCGAGAGTGGCCATCAACAACCGGT 421
Db 657 TCAGTAGTTTATGAGGCAATTAATCTGTGCTGTGAGTCAAGTCCAGCAGCGCGCGCA 716
QY 422 TCAGTAGTTTATGAGGCAATTAATCTGTGCTGTGAGTCAAGTCCAGCAGCGCGCGCA 481
Db 717 CCTGCAACCAAGATGTAATGTATGTAGCCGGAAGGGCGCTCTCCAGCTGTACCGGCA 776
QY 482 CTCTGCAACCAAGATGTAATGTATGTAGCCGGAAGGGCGCTCTCCAGCTGTACCGGCA 539
Db 777 CTGCAACCAATGAGCGCTCAGTGTATCCCAACAGCAGTGTGTCAGGCCGAGCTGC 836
QY 540 CTGCAACCAATGAGCGCTCAGTGTATCCCAACAGCAGTGTGTCAGGCCGAGCTGC 599
Db 837 CGCTGGGATTCAGATCTCTTTCCCAAGGCACTGAGCGCTGC 883
QY 600 AGCGCGGAGTCAAGATCTCTCTCTGAGGCACTGAAACACCGC 646
RESULT 10
LOCUS MMPNEX3 1253 bp DNA ROD 10-SEP-1996
DEFINITION M.musculus gene encoding prepronociceptin, exon 3.
ACCESSION X97372
NID 91532038
KEYWORDS prepronociceptin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 1253)
AUTHORS Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C. and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
MEDLINE 96333281
REFERENCE 2 (bases 1 to 1253)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite Libre de
Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM
FEATURES
source
1..1253
/organism="Mus musculus"
/strain="129/Sv"
/db_xref="taxon:10090"
/clone_id="lambda FIXII (Stratagene #946305)"
605..1087
/number=3
/label=ex3
/usedin=X97371:MMPN_CDS
/usedin=X97371:MMPN_mRNA
/product="prepronociceptin"
BASE COUNT 356 a 307 c 323 g 267 t
ORIGIN
Query Match 32.1%; Score 299; DB 29; Length 1253;
Best Local Similarity 89.8%; Pred. No. 4.21e-159;
Matches 439; Conservative 0; Mismatches 32; Indels 18; Gaps 6;
Db 600 TGCAGACGTGCATCTCCAGAGTGTAGAGAGAGTCTCCCGCGCTGTGACATGTAT 659
QY 83 TGAAGCTGTGATCTCTCAGAGTGTAGAGAGAGTCTCTCTCTGAGACTCTTT 142

```


DB 660 GCACCAAGTCATGGCCAGTGGCTCCGGGAGCTCAGACCTGCTGACCCAGACCTTGTCT 719
 |||||||
 OY 143 GCACCAAGTCATGGCCAGTGGCTCCGGGAGCTCAGACCTGCTGACCCAGACCTTGTCT 202
 |||||||
 DB 720 CACCTGCTCTTTACAGCCGAAAGGCTCGAGATGAGCAGCCTGGAAGAGAAATGCCGGTG 779
 |||||||
 OY 203 CCCTGCTCTTTACAGCCGAAAGGCTCGAGATGAGCAGCCTGGAAGAGAAATGCCGGTG 252
 |||||||
 DB 780 TCCGAGCTTGTGCAAGTGGCAGATGAGCAGCCTGCGGAGATGCTGAGCCTGCGCGAG 839
 |||||||
 OY 263 TCAGAGATGCTGTGCAAGCCGAGAGCAGAGCTGAGGAGATGAGCAGCCTGCTGCGAG 322
 |||||||
 DB 840 ATCTGAGCTGCGCAGATGAGCAGCTGAGAGGTGGAGCAGAAAGCAGCTGAGAAAGT 899
 |||||||
 OY 323 ATG---AG---GC-C---GA---TGAG---TGAGCAGAAAGCAGCTGCAAAAAGT 364
 |||||||
 DB 900 TTGGGGGCTTACCGGGGCGCGAAATCAGCCGGAAGTGGCAGCAAGAGAGGTTCA 959
 |||||||
 OY 365 TTGGGGGCTTACCGGGGCGCGAAATCAGCCGGAAGTGGCAGCAAGAGAGGTTCA 424
 |||||||
 DB 960 GTGAGTTTATGAGCAGTACCTGCTGCTGAGATGAGCAGTCAAGTCAAGCGCGCGACCC 1019
 |||||||
 OY 425 GTGAGTTTATGAGCAGTACCTGCTGCTGAGATGAGCAGTCAAGCAGCGCGCGACTC 484
 |||||||
 DB 1020 TGCACCAAGATGTAATGTGTAGCAGAGAGAGCCCTCCAGCTGACCGCGCACTGCA 1079
 |||||||
 OY 485 TGCACCAAGATGTAATGTGTAGCAGAGAGAGCCCTCCAGCTGACCGCGCACTGCA 544
 |||||||
 DB 1080 ACCCATGAG 1088
 |||||||
 OY 545 ACCCATGAG 553

RESULT 11
 LOCUS AB005251 1209 bp mRNA MAM 09-APR-1998
 DEFINITION Bos taurus mRNA for nociceptin/orphanin FQ precursor, complete cds.
 ACCESSION AB005251
 MID 93041768
 KEYWORDS nociceptin/orphanin FQ precursor.
 SOURCE Bos taurus
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
 1 (sites)
 Okuda-Ashitaka, E., Minami, T., Tachibana, S., Yoshihara, Y.,
 Nishitani, Y., Kimura, T. and Ito, S.
 Nocistatin, a peptide that blocks nociceptin action in pain
 transmission
 Nature 392 (6673), 286-289 (1998)
 JOURNAL Nature 392 (6673), 286-289 (1998)
 MEDLINE 98180619
 REFERENCE 2 (bases 1 to 1209)
 AUTHORS Okuda-Ashitaka, E.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko
 Okuda-Ashitaka, Kansai Medical University, Department of Medical
 Chemistry, 10-15 Fumizono, Moriyasu, Osaka, Japan, Moriyasu,
 Osaka 570, Japan (E-mail:ashitaka@kai.kmu.ac.jp, Tel:06-992-1001,
 Fax:06-992-1781)

FEATURES
 source location/Qualifiers

1..1209
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 210..740
 /product="nociceptin/orphanin FQ precursor"
 /db_xref="PDB:d1026347"
 /db_xref="PDB:g3041769"
 /translation="MKILFCDLLLSLFSSVSSCKDCLVCEKRLPLDPSLEMC
 ILECEKATSPLEPTCTRYKVMAGPADPDAALDPRASEMQLKMPYVR
 SLEFQKTEPGLEVEGEIEQKQKQKRGFTGARKSARKLANQKRFSEFMQYLVLS
 MOSSORRLPLHONGNA"

BASE COUNT 240 a 359 c 341 g 269 t

ORIGIN

Query Match 31.8%; Score 296; DB 20; Length 1209;
 Best Local Similarity 79.8%; Pred. No. 3.11e-157;
 Matches 510; Conservative 0; Mismatches 112; Indels 17; Gaps 10;

DB 250 TCTCCAGCTGTTCACAGCAGTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
 |||||||
 OY 2 TCTCCAGCTGTTCACAGCAGTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
 |||||||
 DB 310 CCAGCTTTCAGAGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
 |||||||
 OY 62 CGGCTCGGAGAGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
 |||||||
 DB 370 CCAGCTTTCAGAGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
 |||||||
 OY 122 CCAGCTTTCAGAGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
 |||||||
 DB 430 CTGCTACCCAGACAGCAGTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
 |||||||
 OY 182 CTGCTATCCAGAGCTTCAGAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
 |||||||
 DB 490 ATGTGAAGGAATGCGCGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 |||||||
 OY 242 ACCTGAAGGAATGCGCGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
 |||||||
 DB 541 CAGAGCCCG-GCCTGAG 594
 |||||||
 OY 302 CAGATCAG 361
 |||||||
 DB 595 GGTTCGGGGGCTTCACCGGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654
 |||||||
 OY 362 GGTTCGGGGGCTTCACCGGGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
 |||||||
 DB 655 TCAGTGTATGAG 714
 |||||||
 OY 422 TCAGTGTATGAG 481
 |||||||
 DB 715 CCCTGACCAAGATGTAATGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
 |||||||
 OY 482 CTCTGACCAAGATGTAATGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
 |||||||
 DB 775 CTTCACCAAGATGTAATGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 834
 |||||||
 OY 540 CTTCACCAAGATGTAATGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
 |||||||
 DB 835 CGGCTGGGATGAG 873
 |||||||
 OY 600 AGGCGGAGATGAG 638
 |||||||

RESULT 12
 LOCUS MMU44027 372 bp mRNA ROD 30-MAY-1996
 DEFINITION Mus musculus orphanin FQ/nociceptin precursor mRNA, partial cds.
 ACCESSION U044027
 MID 91335869
 KEYWORDS house mouse.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
 Murinae; Mus.

REFERENCE 1 (bases 1 to 372)
 AUTHORS Pan, Y.X., Xu, J. and Pasternak, G.W.
 TITLE Cloning and expression of a cDNA encoding a mouse brain orphanin
 FQ/nociceptin precursor
 Biochem. J. 315 (Pt 1), 11-13 (1996)
 JOURNAL Biochem. J. 315 (Pt 1), 11-13 (1996)
 MEDLINE 96207555
 REFERENCE 2 (bases 1 to 372)
 AUTHORS Pan, Y.-X.
 TITLE Direct Submission
 JOURNAL Submitted (28-DEC-1995) Xing-Xian Pan, Neurology, Memorial
 Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY

THIS PAGE BLANK (USPTO)

Q#	T	L	H	Q	N	G	N
RESULT	2						
ID	PNOC.MOUSE	STANDARD;					
AC	064387; 061105; 061938;					PRT;	187 AA.
DT	01-NOV-1997 (REL. 35, CREATED)						
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)						
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)						
DE	NOCICEPTIN PRECURSOR (ORPHANIN FQ) (PNOC) (N23K / N27K).						
DN	PNOC OR NPNC.						
OS	MUS MUSCULUS (MOUSE).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;						
CC	EUTHERIA; RODENTIA.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	TISSUE-BRAIN.						
RX	MEDLINE: 96216718.						
RA	HOUTANI T., NISHI M., TAKESHIMA H., NUKADA T., SUGIMOTO T.;						
RL	BIOCHEM. BIOPHYS. RES. COMMON. 219:714-719(1996).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE: 96106851.						
RA	SAITO Y., MARYAMA K., SAITO T.C., KAWASHIMA S.;						
RL	BIOCHEM. BIOPHYS. RES. COMMON. 217:539-545(1995).						
RN	[3]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN-129/SV;						
RX	MEDLINE: 96323281.						
RA	MOLLEREAU C., SIMONS M.-J., SOULARUE P., LINERS F., VASSART G.,						
RA	MEUNIER J.-C., PARENTIER M.;						
RA	PROC. NATL. ACAD. SCI. U.S.A. 93:8666-8670(1996).						
RN	[4]						
RP	SEQUENCE FROM N.A. (LONG FORM).						
RX	MEDLINE: 96279082.						
RA	SAITO Y., MARYAMA K., KAWANO H., HAGINO-YAMAGISHI K.,						
RA	KAMAMURA K., SAITO T.C., KAWASHIMA S.;						
RL	J. BIOL. CHEM. 271:15615-15622(1996).						
RN	[5]						
RP	SEQUENCE OF 79-187 FROM N.A.						
RC	TISSUE-BRAIN;						
RX	MEDLINE: 96207555.						
RA	PAN Y.-X., XU J., PASTERNAK G.W.;						
RL	BIOCHEM. J. 315:11-13(1996).						
CC	- FUNCTION: NOCICEPTIN IS THE LIGAND OF THE OPIOID RECEPTOR-LIKE						
CC	RECEPTOR (OPEL1). IT MAY ACT AS A TRANSMITTER IN THE BRAIN BY						
CC	MODULATING NOCICEPTIVE AND LOCOMOTOR BEHAVIOR. MAY BE INVOLVED						
CC	IN NEURONAL DIFFERENTIATION AND DEVELOPMENT. WHEN ADMINISTERED						
CC	INTRACEREBROVENTRICULARLY, NOCICEPTIN INDUCES HYPERALGESIA AND						
CC	DECREASES LOCOMOTOR ACTIVITY.						
CC	- SUBCELLULAR LOCATION: SECRETED INTO THE ENDOPLASMIC RETICULUM.						
CC	- TISSUE SPECIFICITY: BRAIN AND SPINAL CORD. LOW LEVELS IN						
CC	KIDNEY AND SPLEEN.						
CC	- DEVELOPMENTAL STAGE: IN EMBRYONIC BRAIN, FIRST DETECTED AT DAY 14						
CC	AND IN POSTNATAL BRAIN, LEVELS INCREASE IN DAY 1 AND DAY 18.						
CC	LEVELS DECREASE SIGNIFICANTLY IN ADULTS.						
CC	- PTR: SPECIFIC ENZYMATIC CLEAVAGES AT PAIRED BASIC RESIDUES						
CC	PROBABLY YIELD OTHER ACTIVE PEPTIDES BESIDES NOCICEPTIN.						
CC	- PTR: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO						
CC	BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.						
CC	- ALTERNATIVE PRODUCTS: A LONGER FORM, N27K IS PROBABLY PRODUCED						
CC	BY ALTERNATIVE SPLICING.						
CC	- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS						
CC	FAMILY.						
DR	EMBL; D82866; D1012281; -						
DR	EMBL; D50056; D1009395; -						
DR	EMBL; X97373; Z244874; -						
DR	EMBL; X97371; Z244782; -						
DR	EMBL; X97372; Z244782; JOINED.						
DR	EMBL; U44027; G1335870; -						
DR	EMBL; D50053; G1304162; -						
DR	MGD; MG1:105308; NPNC1.						
DR	PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.						

DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DE BETA-NEOENDORPHIN-DYNORPHIN PRECURSOR (PROENKEPHALIN B PRECURSOR)
 DE (PREPRODYNORPHIN).
 GN PDYN.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA:
 OC EUTHERIA: RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 90190668.
 RA DOUGLASS J., MCMURRAY C.T., GARRETT J.E., ADELMAN J.P., CALAVETTA L.,
 RL MOL. ENDOCRINOL. 3:2070-2078(1989).
 RN [2]
 RP SEQUENCE OF 45-248 FROM N.A.
 RC TISSUE-HYPOTHALAMUS;
 RX MEDLINE: 85216680.
 RA CIVELLI O., DOUGLASS J., GOLDSTEIN A., HERBERT E.;
 RL PROC. NATL. ACADEM. SCI. U.S.A. 82:4291-4295(1985).
 CC -1- PTM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 CC BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 CC FAMILY.
 DR EMBL: M32784; G204038; -;
 DR EMBL: M32783; G204038; JOINED.
 DR EMBL: M10088; G204040; -;
 DR EMBL: M10088; E10237; ALT_SEQ.
 DR EMBL: M10088; E10239; ALT_SEQ.
 DR EMBL: M10088; E10241; ALT_INIT.
 DR PIR: A01480; DEFTP.
 DR PROSITE: PS01252; OPIOIDS_PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; HYPOTHALAMUS; NEUROPEPTIDE;
 KW ENDOPEPTIN; CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 21
 FT PEPTIDE 166 174 ALPHA-NEOENDORPHIN.
 FT PEPTIDE 166 170 LEU-ENKEPHALIN 1.
 FT PEPTIDE 202 218 DYNORPHIN A.
 FT PEPTIDE 202 206 LEU-ENKEPHALIN 2.
 FT PEPTIDE 221 225 DYNORPHIN B.
 FT PEPTIDE 221 225 LEU-ENKEPHALIN 3.
 FT CONFLICT 181 181 T -> A (IN REF. 2).
 SO SEQUENCE 248 AA: 28078 MW: 6A646802 CRC32;
 Query Match 3.88; Score 175; DB 1; Length 248;
 Best Local Similarity 40.0%; Pred. No. 1,32e-05;
 Matches 46; Conservative 20; Mismatches 49; Indels 0; Gaps 0;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86011328.
 RA VAN HET SCHIP A.D., SANALLO J., BROOS J., OPHUIS J., MOUET M.,
 RA GRUBER M., AB G.;
 RL J. MOL. BIOL. 196:245-260(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 86058863.
 RA NARDELLI D., HET SCHIP F.D., GERBER-HUBER S., HAERLIGER J.A.,
 RA GRUBER M., AB G., WAHLI W.;
 RL J. BIOL. CHEM. 262:15377-15385(1987).
 RN [3]
 RP SEQUENCE OF 1-71 FROM N.A.
 RX MEDLINE: 84131940.
 RA WALKER P., BROWN-LUEDDI M., GERMOND J.E., WAHLI W.,
 RA WEILINK F.C.P.W., VAN HET SCHIP A.D., ROELINK H., GRUBER M.,
 RA GEERT A.B.;
 RL EMBO J. 2:2271-2279(1983).
 RN [4]
 RP SEQUENCE OF 1-71 FROM N.A.
 RX MEDLINE: 84118805.
 RA BURCH J.B.E.;
 RL NUCLEIC ACIDS RES. 12:1117-1135(1984).
 RN [5]
 RP SEQUENCE OF 1092-1338 FROM N.A.
 RX MEDLINE: 85023316.
 RA BYRNE B.M., VAN HET SCHIP A.D., VAN DE KLUNDERT J.A.M., ARNBERG A.C.,
 RA GRUBER M., AB G.;
 RL BIOCHEMISTRY 23:4275-4279(1984).
 RN [6]
 RP SEQUENCE OF 1112-1188.
 RX MEDLINE: 86056531.
 RA CLARK R.C.;
 RL INT. J. BIOCHEM. 17:983-988(1985).
 RN [7]
 RP SEQUENCE OF 1567-1580, AND IDENTIFICATION OF CHAINS.
 RC TISSUE-LIVER;
 RX MEDLINE: 95322425.
 RA YAMAMURA J.-I., ADACHI T., AOKI N., NAKAJIMA H., NAKAMURA R.,
 RA MATSUDA T.;
 RL BIOCHIM. BIOPHYS. ACTA 1244:384-394(1995).
 CC -1- FUNCTION: PRECURSOR OF THE MAJOR EGG-YOLK PROTEINS THAT ARE
 CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS
 CC ORGANISMS.
 CC -1- FUNCTION: PHOSVITIN IS BELIEVED TO BE OF IMPORTANCE IN
 CC SEQUESTERING CALCIUM, IRON AND OTHER CATIONS FOR THE DEVELOPING
 CC EMBRYO.
 CC -1- TISSUE SPECIFICITY: AFTER INCORPORATION FROM SERUM VIA A SPECIFIC
 CC RECEPTOR, IT IS CLEAVED INTO FOUR FRAGMENTS, HEAVY AND LIGHT CHAIN
 CC LIPOVITELLIN, PHOSHOVITIN AND YGP40, AND YGP40 IS RELEASED INTO
 CC THE YOLK PLASMA BEFORE OR DURING COMPARTMENTATION OF LIPOVITELLIN-
 CC PHOSVITIN COMPLEX INTO THE YOLK GRANULE.
 CC -1- INDUCTION: VITELLOGENIN IS SYNTHESIZED IN THE LIVER OF OVIPAROUS
 CC VERTEBRATES IN RESPONSE TO STEROID (ESTROGEN) INDUCTION.
 CC -1- VITELLOGENIN II IS THE MOST ABUNDANT OF THE THREE VITELLOGENINS
 CC (I, II, AND III).
 CC -1- PTM: PHOSVITIN, AN EGG YOLK STORAGE PROTEIN, IS ONE OF THE MOST
 CC HIGHLY PHOSPHORYLATED (10%) PROTEINS IN NATURE.
 CC -1- PTM: CATHEPSIN D IS RESPONSIBLE FOR INTRACYTIC PROCESSING OF
 CC VITELLOGENIN.
 CC -1- PTM: MAY CONTAIN INTRAMOLECULAR DISULFIDE BONDS.
 DR EMBL: X00345; G63873; -;
 DR EMBL: M18060; G212881; -;
 DR EMBL: X00204; G63885; -;
 DR EMBL: X13607; G63887; -;
 DR EMBL: K02113; G212879; -;
 RL PIR: A23177; VTCB2.
 KW GLYCOPROTEIN; PHOSPHORYLATION; STORAGE PROTEIN; SIGNAL.
 FT SIGNAL 1 15
 FT CHAIN 16 1850 VITELLOGENIN II.
 FT CHAIN 16 1111 LIPOVITELLIN I.
 FT CHAIN 1112 1328 PHOSVITIN.

NATURE 306:611-614(1983).
 [2] STRUCTURE BY NMR OF DYNORPHIN.
 RX MEDLINE; 97199318.
 RA TESSMER M.R., KALLICK D.A.;
 RL BIOCHEMISTRY 36:1971-1981(1997).
 CC -1- PM: THE N-TERMINAL CONTAINS 6 CONSERVED CYSTEINES THOUGHT TO
 BE INVOLVED IN DISULFIDE BONDING AND/OR PROCESSING.
 CC -1- SIMILARITY: BELONGS TO THE OPIOIDS NEUROPEPTIDES PRECURSORS
 FAMILY.
 CC EMBL; X02536; -: NOT_ANNOTATED_CDS.
 DR EMBL; K02267; -: NOT_ANNOTATED_CDS.
 DR EMBL; K02268; G182105; -.
 DR EMBL; X00176; G182105; JOINED.
 DR EMBL; X00176; G31155; -.
 DR PIR; A01478; DFHU.
 DR MIM; 131340; -.
 DR PROSITE; PS01252; OPIOIDS_PRECURSOR; 1.
 KW OPIOID PEPTIDE; NEUROTRANSMITTER; HYPOTHALAMUS; NEUROPEPTIDE;
 KM ENDORPHIN; CLEAVAGE ON PAIR OF BASIC RESIDUES; SIGNAL.
 FT SIGNAL 1 20
 FT PEPTIDE 175 183 BETA-NEOENDORPHIN.
 FT PEPTIDE 207 223 DYNORPHIN.
 FT PEPTIDE 226 230 LEU-ENKEPHALIN.
 FT PEPTIDE 226 238 RIMORPHIN.
 FT PEPTIDE 226 254 LETMORPHIN.
 SQ SEQUENCE 254 AA; 28385 MW; 49A11156 CRC32;

Query Match 3.5%; Score 165; DB 1; Length 254;
 Best Local Similarity 38.9%; Pred. No. 3.12e-04;
 Matches 49; Conservative 20; Mismatches 57; Indels 0; Gaps 0;

DB G E G D G D S M G H E D L Y K R Y G G F
 Dt 569 CNGNGARGNGNGAYGNGAWSNATGNGCARGAYTNTNTAARMGTAVGNGNT 628
 QY 314 CTGTCCGAGTGGCCGAGTGGAGCAGACGAGCAGCAGTTCAGTGGGCT 373
 Qc V A D E A D E V E Q K Q L Q K R F G G F
 DB L R R I R P K L K W D N O K R Y G G F L
 Dt 629 TTYTNGNNGNATMGNCMAARTNARIGGAYATCARARMNTAVGNGNTTY 688
 QY 374 TCACCTGGGCGGAGTGGCCGAGTGGCCACGAGCAGTTCAGTGGGCT 433
 Qc T G A R K S A R K L A N Q K R F S E F M
 DB R
 Dt 689 TNGNM 694
 QY 434 TGAGGC 439
 Qc R

RESULT 12
 ID P2CG_MOUSE STANDARD; PRT: 420 AA.
 AC 061074;
 Dt 15-JUL-1998 (REL. 36, CREATED)
 Dt 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 Dt 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROTEIN PHOSPHATASE 2C GAMMA ISOFORM (EC 3.1.3.16) (PP2C-GAMMA)
 DE (FIBROBLAST GROWTH FACTOR INDUCIBLE PROTEIN 13) (FIN13) (FRAGMENT).
 GN PPM1C OR FIN13.
 OS MUS MUSCULUS (MOUSE).
 CC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; ROSENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96226356.
 RA GUTHRIE M.A., SELDIN M., BASILICO C.;
 RL ONCOGENE 12:1267-1278(1996).
 CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
 ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC -1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. LOW LEVEL OF

CC EXPRESSION IN KIDNEY.
 CC -1- INDUCTION: BY FIBROBLAST GROWTH FACTOR.
 CC -1- SIMILARITY: BELONGS TO THE P2C FAMILY.
 DR EMBL; U42383; G1353705; ALT_INIT.
 DR MGD; MG1:106065; FIN13.
 DR PROSITE; PS01032; PP2C; PARTIAL.
 KW HYDROLASE; MAGNESIUM; MANGANESE; MULTIGENE FAMILY.
 FT NON_TER 1 1
 FT METAL 316 316 MANGANESE 2 (BY SIMILARITY).
 FT METAL 371 371 MANGANESE 2 (BY SIMILARITY).
 SQ SEQUENCE 420 AA; 45375 MW; D50BC19 CRC32;

Query Match 3.5%; Score 161; DB 1; Length 420;
 Best Local Similarity 39.1%; Pred. No. 1.07e-03;
 Matches 52; Conservative 31; Mismatches 48; Indels 2; Gaps 2;

DB D E V E E E D S E E C S E D E D G
 Dt 425 SNGAYGARTNGARARGARGAYGAYWSNGARGARTGWSNG-ARGAYGARGAGN 483
 QY 232 GAGATGACAGACCTGAGAGAGATGCCGCTGTCGAGAGTGTGTCAGCCGAGACCA 291
 Qc D A A P E N A A C Q E C G A A R D A
 DB Y S S E E A E N E E D E D T E E A E E
 Dt 484 TAYWSNNGARGARGCNGARAAYGARGARGAYGAYGACNGARGARGARGARG 543
 QY 292 GAGCTGAGGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 350
 Qc E P E A D A E P V A D E A D E V E X E A

DB D D D E
 Dt 544 GAYGAYGAYGARG 556
 QY 351 GCTGCGAGAAAGG 363
 Qc A A E K

RESULT 13
 ID P2CG_BOVIN STANDARD; PRT: 543 AA.
 AC P79126;
 Dt 15-JUL-1998 (REL. 36, CREATED)
 Dt 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 Dt 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROTEIN PHOSPHATASE 2C GAMMA ISOFORM (EC 3.1.3.16) (PP2C-GAMMA)
 DE (MAGNESIUM-DEPENDENT CALCIUM INHIBITABLE PHOSPHATASE) (MCP).
 GN PPM1C.
 OS BOS TAURUS (BOVINE).
 CC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; ARTIODACTYLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX HDANG C.T., QIN K.;
 RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
 ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
 CC -1- COFACTOR: BINDS TWO MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 DR EMBL; U01159; G1754694; -.
 DR PROSITE; PS01032; PP2C; 1.
 KW HYDROLASE; MAGNESIUM; MANGANESE; MULTIGENE FAMILY.
 FT METAL 40 40 MANGANESE 1 (BY SIMILARITY).
 FT METAL 41 41 MANGANESE 1 (BY SIMILARITY).
 FT METAL 60 60 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 439 439 MANGANESE 2 (BY SIMILARITY).
 FT METAL 494 494 MANGANESE 2 (BY SIMILARITY).
 FT DOMAIN 258 324 ASP/GLU-RICH ACIDIC.
 FT DOMAIN 266 272 POLY-GLU.
 FT DOMAIN 306 309 POLY-GLU.
 SQ SEQUENCE 543 AA; 58708 MW; 80B0B0E2 CRC32;

Query Match 3.5%; Score 161; DB 1; Length 543;
 Best Local Similarity 39.8%; Pred. No. 1.07e-03;
 Matches 53; Conservative 30; Mismatches 46; Indels 2; Gaps 2;

FT	REPEAT	74	81	2.
FT	REPEAT	82	89	3.
FT	REPEAT	90	97	4.
FT	REPEAT	97	103	5.
FT	REPEAT	104	111	6.
FT	REPEAT	119	126	7.
FT	REPEAT	127	134	8.
SQ	SEQUENCE	712 AA;	77016 MW;	442EA9F0 CRC32;
Query Match				
Best Local Similarity		3.5%;	Score 165;	DB 1; Length 712;
Matches		30; Conservative	Pred. No. 3.12e-04;	Mismatches 16; Indels 0; Gaps 0;

[illegible]

Search completed: Fri Apr 16 14:46:16 1999
Job time : 276 secs.

[illegible][illegible]

OY	370	GGCTTCACATGGGGCCCGGAGATCAGCCCGGAAAGTTGGCCACACGAAGCGTTACGTGAG	429
OT		G F T T G A R K S A R K L A N Q K R F S E	
Db		F L R	
Dc	670	TTTYYTNMGNM	679
OY	430	:	
OT		F M R	

RESULT	3		
ID	035852	PRELIMINARY;	PRT: 248 AA.
AC	035852;		
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)	
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)	
DT	01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)	
DE	PEREGRYNORPHIN	(FRAGMENT).	
DI	MUS MUSCULUS	(MOUSE).	
OS	EUARCHYOTA: METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;		
OC	EUTHERIA; ROSENTIA.		
CC	[1]		
RP	SEQUENCE FROM N.A.		
KC	STRAIN=129SV; TISSUE=LIVER;		
RA	SHARIFI N., AMENT M., HOCHGESCHWENDER U.;		
RL	SUBMITTED (JUL-1996) TO EMBL/ENBIANK/DBJ DATA BANKS.		
DR	EMBL; U64968; G2341058. -.		
FT	PROSITE; PS01252; OXIDIS_PECURSOR; 1.		
SO	NON TER 248		
SEQUENCE	248 AA; 18009 MW; 793AEFA1 CRC32;		

Query Match	3.9%	Score 180;	DB 11;	Length 248;
Best Local Similarity	39.28;	Pred. No. 7.46e-06;		
Matches	51; Conservative	22; Mismatches	57; Indels	0; Gaps 0;

Dd		E D G G Q D G D Q V G H E D L Y K R Y G	
Dt	550	GARGAYGNGNCARCAVGNCACBRTNGNCAYGARGAYTNTIYAAEMGTATVGN	609
Qy	310	GAECCTTCGCAGATGAGCCGATGAGGTGGAGCAGAAGCACGTCGACGAAGATTGGG	369
Qt		E P V A D E A D E V E Q K Q L Q K R F G	
Dd		G F L R R I R P K L K W D N O Q K R Y G G	
Dt	610	GGNTTYTNMGMMGATIMGMCCNAARTNAPRGGGATYAAYCARARMGTATVANGNGN	668
Qy	370	GGCTTCACCTGGGCCCGGAGACTCAGCCCGGGAAGTTGGCCAACAGGAAGCGGTCACGTGAG	429
Qt		G F T G A R K S A R K L A N Q K R F S E	
Dd		F L R	
Dt	670	TYYTNMGMN 679	
Qy	430	TTTTATGAGGC 439	
Qt		F M R	
RESULT	4		
ID	Q69564	PRELIMINARY;	PRT; 715 AA.
AC	Q69564;		
Dt	01-NOV-1996	(TREMBL.REL. 01, CREATED)	
DT	01-NOV-1996	(TREMBL.REL. 01, LAST SEQUENCE UPDATE)	
Dt	01-AUG-1998	(TREMBL.REL. 07, LAST ANNOTATION UPDATE)	
DE	U86, IE2.		
CN	U86.		
OS	HUMAN HERPESVIRUS-6.		
OC	VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESYTRIDAE; BETAHERPESYTRINAE.		
RN	[1]	SEQUENCE FROM N.A.	
RP	STRAIN-U1102, VARIANT A;		
RC	MEDLINE; 95266321.		
KX	GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J., RA MARTIN M.E., ESTATHIOU S., CRAXTON M., MACAULAY H.A.; RL VIROLOGY 209:29-51(1995).		
NP	[2]	SEQUENCE FROM N.A.	
RC	STRAIN-U1102, VARIANT A;		

RX MEDLINE: 90080132.
 RA LAWRENCE G.L., CHEE M., CRAXTON M.A., GOMPELS U.A., HONESS R.W.,
 RA BARRELL B.G.;
 RL J. VIROL. 64:287-299(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 91237802.
 RA CHANG C.K., BALACHANDRAN N.;
 RL J. VIROL. 65:2884-2894(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 91333007.
 RA TEO I.A., GRIFFIN B.E., JONES M.D.;
 RL J. VIROL. 65:4670-4680(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 91226542.
 RA THOMSON B.J., ESTATHIOU S., HONESS R.W.;
 RL NATURE 351:78-80(1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 91374590.
 RA MARTIN M.E.D., NICHOLAS J., THOMSON B.J., NEWMAN C., HONESS R.W.;
 RL J. VIROL. 65:5381-5390(1991).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 92333249.
 RA ESTATHIOU S., LAWRENCE G.L., BROWN C.M., BARRELL B.G.;
 RL J. GEN. VIROL. 73:1661-1671(1992).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 92148942.
 RA GENG Y., CHANDRAN B., JOSEPHS S.F., WOOD C.;
 RL J. VIROL. 66:1564-1570(1992).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93091236.
 RA GOMPELS U.A., CARSS A.L., SUN N., ARRAND J.R.;
 RL DNA SEQ. 3:25-39(1992).
 RN [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 92260671.
 RA NEIPEL F., ELLINGER K., FLECKENSTEIN B.;
 RL J. VIROL. 66:3918-3924(1992).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 92333248.
 RA THOMSON B.J., HONESS R.W.;
 RL J. GEN. VIROL. 73:1649-1660(1992).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93187613.
 RA ELLINGER K., NEIPEL F., FOA-TOMASI L., CAMPADELLI-FIUME G.,
 RA FLECKENSTEIN B.;
 RL J. GEN. VIROL. 74:495-500(1993).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93324882.
 RA GOMPELS U.A., CARRIGAN D.R., CARSS A.L., ARNO J.;
 RL J. GEN. VIROL. 74:613-622(1993).
 RN [14]
 RP SEQUENCE FROM N.A.

RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93389439.
 RA LIU D.X., GOMPELS U.A., NICHOLAS J., LELLIOTT C.;
 RL J. GEN. VIROL. 74:1847-1857(1993).
 RN [15]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94025558.
 RA LIU D.X., GOMPELS U.A., FOA-TOMASI L., CAMPADELLI-FIUME G.;
 RL VIROLOGY 197:12-22(1993).
 RN [16]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93331710.
 RA PELLET P.E., SANCHEZ-MARTINEZ D., DOMINGUEZ G., BLACK J.B., ANTON E.,
 RA GREENAMOVER C., DAMBAUGH T.R.;
 RL VIROLOGY 195:521-531(1993).
 RN [17]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 93323202.
 RA PEIFFER B., BERNEMAN Z.N., NEIPEL F., CHANG C.K., TIRWATNAPONG S.,
 RA CHANDRAN B.;
 RL J. VIROL. 67:4611-4620(1993).
 RN [18]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 95146989.
 RA GOMPELS U.A., MACAULAY H.A.;
 RL J. GEN. VIROL. 76:451-458(1995).
 RN [19]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94047392.
 RA DEMHURST S., DOLIARD S.C., PELLETT P.E., DAMBAUGH T.R.;
 RL J. VIROL. 67:7680-7683(1993).
 RN [20]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX NICHOLAS J.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [21]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94118404.
 RA NICHOLAS J., MARTIN M.;
 RL J. VIROL. 68:597-610(1994).
 RN [22]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94202284.
 RA SCHIEME U., NEIPEL F., SCHREINER D., FLECKENSTEIN B.;
 RL J. VIROL. 68:2978-2985(1994).
 RN [23]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94181269.
 RA THOMPSON J., CHOUDHURY S., KASHANCHI F., DONIGER J., BERNEMAN Z.,
 RA FRENKEL N., ROSENTHAL L.J.;
 RL ONCOGENE 9:1167-1175(1994).
 RN [24]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94167865.
 RA ZHOU Y., CHANG C.K., QIAN G., CHANDRAN B., WOOD C.;
 RL VIROLOGY 199:311-322(1994).
 RN [25]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102, VARIANT A;
 RX MEDLINE: 94202288.
 RA THOMSON B.J., DEMHURST S., GRAY D.;
 RL J. VIROL. 68:3007-3014(1994).
 RN [26]


```
Dt 2045 CMMNSMNAARGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2104
Oy 301 GCAGTGCAGAGCTGCTGCGATGAGCGCGATGAG-GTGAGAGCAGAGCAGCGAA 359
Qt R C R A C R R * G R * X W S R S S C C R R K

Db S S R A S G K A S S E A S S R A S S R N
Dt 2105 CMMNSMNAARGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2164
Oy 360 AACGTTGGGGCTTCTGCTGGGGCGCGAGAGTGGCCCAACGAGAGCG 419
Qt G L G A S L G P G S Q P G S W P T R S G

Db S S R A S S R A S S R A S S R D
Dt 2165 AYSMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2215
Oy 420 GTTCAGTGAATTATGAGGAGTACCTGCTGAGCATGAGTCAAGCCA 470
Qt S V S L * G S T W S * A C S Q A

RESULT 6 PRELIMINARY: PRT: 1500 AA.
ID 042080;
AC 042080;
Dt 01-JAN-1998 (TREMBLREL. 05, CREATED)
Dt 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
Dt 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE IE-2 PROTEIN.
GN IE2HOM.
OS HUMAN HERPESVIRUS 6.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; ROSEOLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UGANDA-1102.
RA SOEREL A., SCHIEWE U., FLECKENSTEIN B., NEIPEL F.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA SOEREL A., SCHIEWE U., FLECKENSTEIN B., NEIPEL F.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF015298; G2384715; -.
DR EMBL: AF015297; G2343283; -.
SO SEQUENCE 1500 AA; 168678 MW; E5F695F8 CRC32;

Query Match 3.98; Score 181; DB 14; Length 1500;
Best Local Similarity 14.28; Pred. No. 5.42e-06;
Matches 67; Conservative 154; Mismatches 243; Indels 7; Gaps 7;

Db S R A S S R D S S R A S S R A S S R A S
Dt 2185 MSNMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2244
Oy 2 TCTCCAGCGTGTAGCAGCTGCTCCAGAGCTGCTCAGAGAGAGGCTCCACC 61
Qt S P A C S A A V P R T A S P A R R G S T

Db S R D S S R A S S R A S S R D S S R A S
Dt 2245 MSNMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2304
Oy 62 CGGCTCGGCGAGCTTAACCTGAAGCTGTGCTCCTGCGAGAGAGGCTTCC 121
Qt R L R A A S T * S C A S S S V K R R S S

Db S R A S S R A S S K A S S S R A S S R A S
Dt 2305 MSNMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2363
Oy 122 -CCGCCCTCTGAGCTCTTGTGACCAAGCCATGGCCGAGCTGAGCAGCTTCC 180
Qt X R P L W T L C T K A M A S D S S S A

Db S R A S S R D S S R A S S K A S S R A
Dt 2364 N-MSNMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2420
Oy 181 CCGCTATCAGAGCTCAGCTCGCTGCTTTACCAAGTCAAGAGCTGAGATCAG 240
Qt C * S R A H V R C L Y S R K P R C S

Db S S R D S S R A S S R A S S R A S S R A
Dt 2421 MSNMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2479
Oy 181 CCGCTATCAGAGCTCAGCTCGCTGCTTTACCAAGTCAAGAGCTGAGATCAG 240
Qt C * S R A H V R C L Y S R K P R C S
```

```
Oy 241 CACGTGAAGATGCGCGCTGTGAGAGTGTGTCGACACCGCAGAGCCTGAG 300
Qt T * R E C R V S G C G A S P R R A * G

Db S S R A S S R A S S R A S S R D S S R A
Dt 2480 CMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2539
Oy 301 GCAGTGCAGAGCTGCTGCGATGAGCGCGATGAG-GTGAGACAGAGAGCTCAGAA 359
Qt R C R A C R R * G R * X W S R S S C C R K

Db S S R A S S R A S S R D S S R A S S R D
Dt 2540 CMMNSMNAARGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2599
Oy 360 AACGTTGGGGCTTCTGCTGGGGCGCGAGAGTGGCCCAACGAGAGCG 419
Qt G L G A S L G P G S Q P G S W P T R S G

Db S S R D S S R A S S R A S S R D
Dt 2600 AYSMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGCMMNSMNGC 2650
Oy 420 GTTCAGTGAATTATGAGGAGTACCTGCTGAGCATGAGTCAAGCCA 470
Qt S V S L * G S T W S * A C S Q A

RESULT 7 PRELIMINARY: PRT: 730 AA.
ID 006634;
AC 006634;
Dt 01-NOV-1996 (TREMBLREL. 01, CREATED)
Dt 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
Dt 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN BICP4 (FRAGMENT).
GN BICP4.
OS BOVINE HERPESVIRUS TYPE 1 (STRAIN JURA (SUBTYPE 1.1));
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94025583.
RA SCHWYZER M., VICER C., MENEKSE O., FRAEFEL C., PACES V.;
RL VIROLOGY 197:349-357(1993).
DR EMBL: L14321; G291538; -.
KW TRANSCRIPTION REGULATION; TRANS-ACTING FACTOR; ACTIVATOR; ZINC-FINGER;
FT DNA-BINDING; EARLY PROTEIN; REPRESSOR.
FT NON TER 730
SO SEQUENCE 730 AA; 72965 MW; 61E83E42 CRC32;

Query Match 3.88; Score 175; DB 14; Length 730;
Best Local Similarity 43.48; Pred. No. 3.62e-05;
Matches 36; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

Db A E D E A E D G E D E A E E E E
Dt 2108 ARGCGNARGAARGCAGARGAARGAARGAARGAARGAARGAARGAARGAARGAARGAARG 2167
Oy 278 AACGCCGAGAGCGAGAGCTGAGAGAGCTGCGAGATGAGAGCGCATGAGG 337
Qt A R D A E P E A D A E P V A D E A D E V

Db E E K E E E
Dt 2168 ARGARGAARGAARGAARGAARGAARGAARGAARGAARGAARGAARGAARGAARGAARG 2190
Oy 338 TGGAGCAGAGAGCTGACAGAA 360
Qt E Q K Q L Q K

RESULT 8 PRELIMINARY: PRT: 1482 AA.
ID 014550;
AC 014550;
Dt 01-NOV-1996 (TREMBLREL. 01, CREATED)
Dt 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
Dt 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE ZINC-FINGER DNA-BINDING PROTEIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
```

[illegible]

Db 738 E E E E E E Y
RGARGARGARGARGARTAYG 760
OY 345 GAAGCAGCTGCGAAGAGTTTG 367
Ot K O L O K R F

RESULT 12
ID 057149 PRELIMINARY; PRT; 949 AA.
AC 057149;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE SIMILAR TO H166A U86.
GN H87.
OS HUMAN HERPESVIRUS 6.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; BETAHERPESVIRINAE;
OC ROSEOLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HST;
RA KOSUGE H., ISEGAWA Y., YAMANISHI K.;
RL VIRUS RES. 52:1-14(1997).
DP EMBL: U92288; G2769724;
SO SEQUENCE 949 AA; 104665 MW; 8647921E CRC32;

Query Match 3.68; Score 170; DB 14; Length 949;
Best Local Similarity 15.8%; Pred. No. 1.70e-04;
Matches 70; Conservative 134; Mismatches 233; Indels 7; Gaps 7;

Db 2227 R A S S R A S S R A S S R A S S R A S S R A S S
MGNCNNSNNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2286
OY 29 AGAGCTCCCTCAGCAGAGAGGCTCCACCGCTCCGCGCAGCTTGAAGC 88
Ot R T A S P A R R G S T R L R A A S T * S

Db 2287 R A S S R A S S K T S S R A S S R A S S R A F S
MGNCNNSNNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2346
OY 89 TGTCATCCCTCAGTGTGAAGAGAGGCTTCC-CGCCCTCTGTGACCTTTCGCC 147
Ot C A S S S V K R R S S x R P L W T L C T

Db K A S S S R A S S S R A S S S R A S S S R A S
AARCNNSNNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2403
OY 148 AAAGCCATGGCAGCTGACCTGAGCAGCTCCGCTGCTGATCCAGAGCTTCCGCT 207
Ot K A M A S D S S S S A C C T * S R A H V A

Db S R A S S R A S S S R A S S S R A S S S R A S
WSNNGNG-CNNSNNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNS 2462
OY 208 GCTCTTTACAGTGGAAAGCCTCGAGATGCGACCAAGCAATGCCGCTGTGAGG 267
Ot A L Y S R K P R R C S T * R E C R V S G

Db G R A S S K A S S S G V S S K A S S S R A S
GNGMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2521
OY 268 AGTGTGTCAGAGCCGAGAGCAGCAGCTGAGCAGTGCAGAGCTTCCAGAGCT 327
Ot V G A S S P R R R A * G R C R A C R R * G

Db R R A S S R A S S S R A S S S R A S S S R A S
SNNMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2581
OY 328 GCCCATAG-GTGAGAGAGAGCAGCTGCGAAGAGTTTGCGGCTTCACTGGGCGCG 386
Ot R * X W S R S S C R K G L G A S L G P G

Db S R A S S R A S S S R A S S S R A S S S R A S
SNNMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2641
OY 387 GAACTGAGCCCGGGAAGTTGGCAACCAAGCGTTTCAGTATTTATGAGCGAGTACCT 446
Ot S Q P G S W P T R S G S V S L * G S T W

Db S R A S S R A S

Dt 2642 SNNMNGCNSMNSMNGCNSMNSMNGCNSMNSMNGCNSMNSN 2565
OY 447 GGCTGAGCATGCGAGTCAAGCCA 470
Ot S * A C S O A

RESULT 13
ID 064318 PRELIMINARY; PRT; 1117 AA.
AC 064318; 062519;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR 8 (MEB1) (DELTA E11).
GN TCF8.
OS MOS MUSCULUS (MOUSE).
OC EOKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6; TISSUE-EMBRYO;
RA SERIDO R., TAKAGI T., MORIBE H., YAMAMURA M., HIGASHI Y., KONDOH H.;
RL GENE 173:227-232(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-SPLEEN;
RA WU Y., MONTIYA G.D., RUBIN S.E., BRODIE S.G., JENKINS N.,
RA WILLIAMS T.M.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 96194821.
RA GENNETTA T., KADESCH T.;
RL GENE 169:289-290(1996).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
IN THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC CRYSTALLIN ENHANCER BINDING FACTOR AND HAMSTER ZINC FINGER PROTEIN
B2P.
DR EMBL: U26259; G833849; -;
DR EMBL: D76432; G1027500; -;
DR EMBL: L48363; G1220420; -;
DR MGD; MGI:98508; TCF8.
DR POSITIVE; P500028; ZINC FINGER C2H2; 5.
KW DNA-BINDING; NUCLEAR PROTEIN; ZINC-FINGER; METAL-BINDING; REPRESSOR;
KW TRANSCRIPTION REGULATION.

FT ZN_FING 152 173
FT ZN_FING 182 202
FT ZN_FING 222 242
FT ZN_FING 250 272
FT ZN_FING 511 534
FT DNA_BIND 560 619
FT ZN_FING 884 904
FT ZN_FING 912 932
FT ZN_FING 940 959
FT CONFLICT 33 33
FT CONFLICT 57 57
FT CONFLICT 125 125
FT CONFLICT 146 146
FT CONFLICT 313 313
FT CONFLICT 353 353
FT CONFLICT 400 400
FT CONFLICT 461 461
FT CONFLICT 528 528
FT CONFLICT 664 664
FT CONFLICT 744 744
FT CONFLICT 822 822
FT CONFLICT 1062 1062
SO SEQUENCE 1117 AA; 122464 MW; 78FB4707 CRC32;

Query Match 3.68; Score 167; DB 11; Length 1117;


```

Db  P A D P E L T S A A L Y Q S K A S E M Q
Dc 181 CCNGNGAYCCNGARNTYACNMGNGNNGNYTAYCARMSNAARCSNBARATGCAR 240
Qy 181 CCGTGCATCCAGCTCGCTCGCTGCTTTACCATCCAAAGCCTCGAGATGCAG 240
Qt  P A D P E L T S A A L Y Q S K A S E M Q

Db  H L K R M P R V R S V V Q A R D A E P E
Dc 241 CAYTNAARMGNATGCCNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGN 300
Qy 241 CACCTGAAGAGATGCCGCTGTCAGAGAGTGTGTCAGAGAGAGAGAGAGAGAGAG 300
Qt  H L K R M P R V R S V V Q A R D A E P E

Db  A D A E P V A D E A D E V E Q K Q L Q K
Dc 301 GNGAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAG 360
Qy 301 GNGAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAG 360
Qt  A D A E P V A D E A D E V E Q K Q L Q K

Db  R F G G F T G A R K S A R K L A N Q K R
Dc 361 MONTYGGGNTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 420
Qy 361 AGTTGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Qt  R F G G F T G A R K S A R K L A N Q K R

Db  F S E F M R Q Y L V L S M Q S S Q R R R
Dc 421 TTYMSNGARTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 480
Qy 421 TTYMSNGARTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 480
Qt  F S E F M R Q Y L V L S M Q S S Q R R R

Db  T L H Q N G N
Dc 481 ACNTNCAVCARAAAGNAYGT 503
Qy 481 ACTCTGACACAGATGTATGT 503
Qt  T L H Q N G N

RESULT 2
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
ACCESSIONS JC6151
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Airdat, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession J06151
#molecule_type mRNA
#residues 1-181 #label NOT
#cross_references GB:U48262; NID:g185011; PID:g185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opioid receptors. It is involved in development.
KEYWORDS
SUMMARY #length 181 #molecular_weight 20478 #checksum 3303

Query Match 42.4%; Score 1978; DB 2; Length 181;
Best Local Similarity 55.1%; Pred. No. 0.00e+00;
Matches 277; Conservative 133; Mismatches 93; Indels 0; Gaps 0;

Db  L S S V F S S C P E D C L T C Q E R L H
Dc 40 YTNWSNMSNGTNTYMSNMSNTGCGNGARGATGYTNCARGARBMNGTNCAY 99
Qy 1 CTCTCCAGCGTGTCCAGCAGCTGTCCGAGAGTCTCTCCAGAGGAGGAGGCTCCAC 60
Qt  L S S V F S S C P E D C L T C Q E R L H

Db  P A P G S F N L K L C I L Q C E E K V F
Dc 100 CCNGCNGCNGSNTNTYTAATYNTGATHTTNCARTGYGARGARATGTTT 159
Qy 61 CCGGCTCGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Qt  P A P G S F N L K L C I L Q C E E K V F

Db  P R P L W T L C T K A M A S D S E Q L S
Dc 160 CCNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 219
Qy 121 CCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
Qt  P R P L W T L C T K A M A S D S E Q L S

Db  P A D P E L T S A A L Y Q S K A S E M Q
Dc 220 CCNGCNGCNGARNTYACNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 279
Qy 181 CCGTGCATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAG 240
Qt  P A D P E L T S A A L Y Q S K A S E M Q

Db  H L K R M P R V R S V V Q A R D A E P E
Dc 280 CAYTNAARMGNATGCCNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGN 339
Qy 241 CACCTGAAGAGATGCCGCTGTCAGAGAGTGTGTCAGAGAGAGAGAGAGAGAG 300
Qt  H L K R M P R V R S V V Q A R D A E P E

Db  A D A E P V A D E A D E V E Q K Q L Q K
Dc 340 GNGAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAG 399
Qy 301 GNGAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAGATCCAG 360
Qt  A D A E P V A D E A D E V E Q K Q L Q K

Db  R F G G F T G A R K S A R K L A N Q K R
Dc 400 MONTYGGGNTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 459
Qy 361 AGTTGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Qt  R F G G F T G A R K S A R K L A N Q K R

Db  F S E F M R Q Y L V L S M Q S S Q R R R
Dc 460 TTYMSNGARTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 519
Qy 421 TTYMSNGARTTATGNCNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNGNMGNG 480
Qt  F S E F M R Q Y L V L S M Q S S Q R R R

Db  T L H Q N G N
Dc 520 ACNTNCAVCARAAAGNAYGT 542
Qy 481 ACTCTGACACAGATGTATGT 503
Qt  T L H Q N G N

RESULT 3
ENTRY JC4502 #type complete
TITLE opioid-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 10-Sep-1997
ACCESSIONS JC4502; JC4552
REFERENCE JC4502
#authors Saito, Y.; Maruyama, K.; Saïdo, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 #label SNI
#cross_references DBD:050056; NID:g1857023; PID:01009395; PID:g1217906
#experimental_source NS20Y neuroblastoma cells
REFERENCE JC4552
#authors Houtanli, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto, T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719

```


#title	Structure and regional distribution of nociceptin/orphanin Fy precursor
#accession	JC4652
#molecule_type	mRNA
#residues	1-187
#cross-references	DBS:DB82866; NID:g1311472; PID:d1012281; PID:g1311473
#experimental_source	brain
#note	The authors translated the codon CTC for residue 72 as Lys and GAC for residue 76 as Glu.
COMMENT	This protein belongs to the opiate precursor family, and functions as a neuropeptide precursor and plays a role in neuronal differentiation and development.
GENETICS	
#gene	N23K
KEYWORDS	brain; differentiation; G protein-coupled receptor; neuropeptide
FEATURE	
1-18	#domain signal region #status predicted #label SIG
19-187	#product neuropeptide precursor, N23K #status predicted #label Mtr\
109-126	#region D-E-E-P-G-A motif\
141-157	#product neuropeptide #status predicted #label NRP\
160-176	#product heptadecapeptide #status predicted #label HPP
SUMMARY	#length 187 #molecular-weight 20864 #checksum 8758
Query Match	27.4%: Score 1276; DB 2; Length 187;
Best Local Similarity	51.8%: Pred. No. 4.77e-200;
Matches	270; Conservative 123; Mismatches 110; Indels 18; Gaps 8;
Db	L S S V F S S C P R D C L T C O E K L H
Dt	YTNNSNGTNTTYSNWSNTGTCGNGMGAATGYTNACNTGYCARARARYNCAY 99
Qy	1 CTCCTCAGCGGTGTACACAGCTGCCGAGACAGTGCCTCAGTCCGAGAGGCTCCAC 60
Ot	L S S V F S S C P E D C L T C O E K V F
Db	P A P D S F N L K T C I L O C E E K V F
Dt	100 CCNGCNCNGMYNSNTTYAAATYTNARACTGTATHTYTCARTGTGARCARARNTTY 159
Qy	61 CCGGCTCGGGACCTTCAACCTGAACCTGTGATCTCCAGTGTGAAGAGGTTCTC 120
Ot	P A P G S F N L K L C I L O C E E K V F
Db	P R P L W T V C T K V M A S G S G O L S
Dt	160 CCNGNCNTYNTGACNGTNGTCAACAGTATGAGCMMNSGNCNCAATYTNNSN 219
Qy	121 CCCCCTCTCTGACTCTTGGCAACCAAGCCATGACGAGCTGACGCTCAGC 180
Ot	P R P L W T L T C C T K A M A S D S E Q L S
Db	P A D P E L V S A L Y O P K A S E E M O
Dt	220 CCNGCNGAYCCNGARYNTNGTNSGNCNGYNTYATCARCNAARCCNSNARATGCA 279
Qy	181 CCTCTGTATCAGAGCTCAGCTCGCGTCTTATACAGCGAAACCTCGGAGATGAG 240
Ot	P A D P E L T S A A L Y O S K A S E M Q
Db	H L K R M P R V R S L V Q V R D A E P C
Dt	280 CAYTINARMGNAATGCCNMNGTNGMWSYNTGTCARTGNTNGNAYCCNARCNGSN 339
Qy	241 CACCTGAAGAAAGCCGCGGTGAG--GA---GT---GT---G-G-TGCA-AGCCG-- 284
Ot	H L K R M P R V R V X X V V X X A X P X
Db	A D A E P G A D A E P G A D D A E E V E
Dt	340 GCNGAYGCNGARCCNGGNGCNGAYCAGCNGARCCNGGNCNGAYAYCCNARARATNGAR 399
Qy	285 --AGACGAGAGCCTGAGCGAGATGACAGACCTGTGCGAGATAGGCCCATGAGGGAG 342
Ot	X D A E P E A D A E P V A D E A D E V E
Db	O K O L O K R F G G F T G A R K S A R K
Dt	400 CARARCARRYNCARARMGNTTYGNGNTTYACNGNCNNGNARNSNCGNNGNAR 459
Qy	343 CAGAGCGAGCGCAAAAGGTGTGGGGCTTACTAGTGGCCGGAAGCAGCCGGAAG 402
Ot	O K O L O K R R F G G F T G A R K S A R K
Db	L A N O K R F S E F M R O Y L V L S M O
Dt	460 YTNCGNAAYCARARMGNTTYWSGARTTYATGNCNARAYTNTGNTTNSNATGAC 519

[illegible]

```

OY 302 CAGATGAGAGCTGTGGCAGATGAGGCCCATGATGAGTGGACACAGACAGCTGCAGAAA 361
Qt D R A E A D E V E Q K Q L Q K R

Db F G G F T G A R K S A R K L A N Q K R F
Dt 386 GNTTGGGNGNTTAAACGNCNMGNAARMSNGMNAARCAARABAGNT 445
OY 362 GGTGGGGGCTTACGCGGCCGGAAGTACGCCGGAAGTGGCCACCAAGCCGCT 421
Qt F G G F T G A R K S A R K L A N Q K R F

Db S E F M R Q Y L V L S M Q S S Q R R R T
Dt 446 TYSNGARITTYACNTATYATYNTNGTNSATNCARMSNWSNCAARMNGNA 505
OY 422 TCAGTGGTTATGAGGACCTGCTGCGAGACATGACATGACGACGACGCGCA 481
Qt S E F M R Q Y L V L S M Q S S Q R R R T

Db L H Q N G N
Dt 506 CNYTNCAYCARAAYGNAAYGT 527
OY 482 CTCTGCACCAAGATGTAATGT 503
Qt L H Q N G N

RESULT 5
ENTRY 570631 #type fragment
TITLE orphanin FQ/noiceptin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
10-Sep-1997

ACCESSIONS 570631
REFERENCE 570631
#authors Pan, Y.X.; Xu, J.; Pasternak, G.W.
#journal Biochem. J. (1996) 315:11-13
#title Cloning and expression of a cDNA encoding a mouse brain
#accession orphanin FQ/noiceptin precursor.
#status preliminary
#molecule_type mRNA
#residues 1-109 ##label PAN
#cross-references EMBL:U44027; NID:g1335869; PID:g1335870
SUMMARY #length 109 #checksum 9509

Query Match 18.4%; Score 856; DB 2; Length 109;
Best Local Similarity 50.9%; Pred. No. 2.94e-120;
Matches 137; Conservative 60; Mismatches 72; Indels 0; Gaps 0;

Db M P R V R S L V Q V R D A E P G A D A E
Dt 58 ATCCCNNGNTNMGNSNTNGTNCARGTNGAYGNCNGNGNGAR 117
OY 235 ATCCAGACCTGAAGAAATGCCGCGTGTGAGAGTGTGTGAGAGCCGAGAGCAGAG 294
Qt M Q H L K R M P R V R S V V Q A R D A E

Db P G A D A E P G A D A E V E Q K Q L
Dt 118 CCGNGGNCNAYCNGARCCNGNGNGAYGACNGARCCNGNGNGAYGNGAR 177
OY 295 CCGTGGGACATGAGGCTGTGCGACGATGAGGAGGAGGAGGAGGAGGAGGAGG 354
Qt P E A D A E P V A D A D E A D E V E Q K Q L

Db O K R F G G F T G A R K S A R K L A N Q
Dt 178 CAAAGAGNTTNGNTTACNGNGNGMNAARMSNCGNNAARNTNGNCAAR 237
OY 355 CAAAGAGNTTNGNTTACNGNGNGMNAARMSNCGNNAARNTNGNCAAR 414
Qt Q K R F G G F T G A R K S A R K L A N Q

Db K R F S E F M R Q Y L V L S M Q S S Q R
Dt 238 AARNGTITTSNAGRTTATGACNCAATAYTNGTITTSNAGRTTATGACNCAAR 297
OY 415 AAGCGGTATGAGTGTATGAGGACATGACCTGCTGAGCATGAGTCAACCAACG 474
Qt K R F S E F M R Q Y L V L S M Q S S Q R

Db R R T L H Q N G N
Dt 298 MGNMGNACNTTNCAYCARAAYGNAAYGT 326
OY 475 CGCCGACCTCTGCACCAAGATGTAATGT 503

```

```

Qt R R T L H Q N G N

RESULT 6
ENTRY JC6318 #type complete
TITLE prodynorphin - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 14-May-1998 #sequence_revision 29-May-1998 #text_change
24-Sep-1998

ACCESSIONS JC6318
REFERENCE JC6318
#authors Ujanga, H.; Weesner, G.D.; Malven, P.V.
#journal Gene (1997) 186:279-283
#title cDNA sequence and expression of bovine prodynorphin.
#accession JC6318
#molecule_type mRNA
#residues 1-258 ##label JTA
#cross-references GB:U58500; NID:g1519041; PID:g1519042
#experimental_source brain
COMMENT This protein is the precursor for the dynorphin family of
endogenous opioid peptides.

GENETICS
#gene PRODYN
#superfamily proenkephalin
CLASSIFICATION #length 258 #molecular_weight 28789 #checksum 3422
SUMMARY

Query Match 4.4%; Score 205; DB 2; Length 258;
Best Local Similarity 39.7%; Pred. No. 3.05e-08;
Matches 52; Conservative 23; Mismatches 56; Indels 0; Gaps 0;

Db E A G E G E G E V G H E D L Y K R Y G
Dt 576 RGAAGGNGNGARNGARNGARNGARNGARNGARNGARNGARNGARNGARNGAR 635
OY 309 AGAGCTGTGCGAGATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 368
Qt E P V A D E A D E V E Q K Q K R F G

Db G F L R R I R P K L K W D N Q K R Y G
Dt 636 NGNTTYTNNMGNTNMGNTNMGNTNMGNTNMGNTNMGNTNMGNTNMGNTNMGNT 695
OY 369 GGGCTTCACTGGGCGCCGAGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 428
Qt G F T G A R K S A R K L A N Q K R F S E

Db F L R
Dt 696 NTYYTNNMGNT 706
OY 429 GTTATGAGGC 439
Qt F M R

RESULT 7
ENTRY CGR181 #type fragment
TITLE collagen alpha 1(I) chain - rabbit (fragment)
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 12-Aug-1981 #sequence_revision 12-Aug-1981 #text_change
05-Sep-1997

ACCESSIONS A02856
REFERENCE A02856
#authors Bornstein, P.; Nese, R.
#journal Arch. Biochem. Biophys. (1970) 138:443-450
#title The comparative biochemistry of collagen: the structure of
rabbit skin collagen and its relevance to immunochemical
studies of collagen.
#cross-references MUID:70252720
#accession A02856
#molecule_type protein
#residues 1-53 ##label BOR
#experimental_source skin
#note the compositions of CNB1 and CNB2 were determined
#note we have positioned residues by homology with rat tendon
#note alpha 1(I) chain
#note it is assumed that the epsilon carbon of Lys-7 is
converted to an aldehyde group that is involved in

```

```

#note      cross-linking
           the six prolines at the third position of the tripeptide
           repeating unit (G-X-Y) may be hydroxylated
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
               carboxyl-terminal homology; von Willebrand factor type C
               repeat homology
KEYWORDS      coll'd coll; extracellular matrix; glycoprotein; trimer;
               triple helix
FEATURE       #modified_site allysine (lys) #status predicted
SUMMARY       #length 53 #checksum 1461
Insufficient memory To Reconstruct alignment.

RESULT      8
ENTRY       CGRTIS #type fragments
TITLE       collagen alpha 1(I) chain - rat (fragments)
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change
           19-Oct-1995
ACCESSIONS  A90559; A90552; A90353; A90566; A90357; A90362;
           A90379; A91209; A91385; A02854; A02855
REFERENCE   #authors  Bornstein, P.
           #journal    Biochemistry (1969) 8:63-71
           #title      Comparative sequence studies of rat skin and tendon collagen.
           #note       I. The absence of a short sequence at the amino terminus
           of the skin alpha1 chain.
           #cross-references MUID:69155173
           #contents  CNBR0 and CNBR1
           #accession A90559
           #molecule_type protein
           #residues  1-19 #label B01
           #experimental_source tendon
           #note       sequences from skin and tendon appear to be identical
           the amino-terminal tetrapeptide may be removed by
           limited proteolysis during extraction

REFERENCE   #authors  Kang, A.H.; Bornstein, P.; Piez, K.A.
           #journal    Biochemistry (1967) 6:788-795
           #title      The amino acid sequence of peptides from the cross-linking
           region of rat skin collagen.
           #cross-references MUID:67162268
           #contents  CNBR1
           #accession A90552
           #molecule_type protein
           #residues  5-19 #label KAN
           #experimental_source skin
REFERENCE   #authors  Bornstein, P.
           #journal    J. Biol. Chem. (1967) 242:2572-2574
           #title      The incomplete hydroxylation of individual prolyl residues in
           collagen.
           #cross-references MUID:67165368
           #contents  CNBR2
           #accession A92029
           #molecule_type protein
           #residues  20-55 #label B02
           #experimental_source skin and tendon
REFERENCE   #authors  Butler, W.T.; Ponds, S.L.
           #journal    Biochemistry (1971) 10:2076-2081
           #title      Chemical studies on the cyanogen bromide peptides of rat skin
           collagen. Amino acid sequence of alpha1-CB4.
           #cross-references MUID:71263178
           #contents  CNBR4
           #accession A90353
           #molecule_type protein
           #residues  56-102 #label B01
           #experimental_source skin
REFERENCE   #authors  Butler, W.T.
           #note       A90566
           this region (residues 651-671 above) probably

```

```

#journal    Biochemistry (1970) 9:44-50
#title      Chemical studies on the cyanogen bromide peptides of rat skin
           collagen. The covalent structure of alpha1-CB5, the major
           hexose-containing cyanogen bromide peptide of alpha1.
           #cross-references MUID:70085124
           #contents  CNBR5
           #accession A90566
           #molecule_type protein
           #residues  103-139 #label B02
           #experimental_source skin
REFERENCE   #authors  Balian, G.; Click, E.M.; Bornstein, P.
           #journal    Biochemistry (1971) 10:4470-4478
           #title      Structure of rat skin collagen alpha1-CB8. Amino acid
           sequence of the hydroxylamine-produced fragment HA2.
           #cross-references MUID:72136131
           #contents  CNBR8
           #accession A90357
           #molecule_type protein
           #residues  140-238 #label B01
           #experimental_source skin
REFERENCE   #authors  Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
           #journal    Biochemistry (1972) 11:3798-3806
           #title      Structure of rat skin collagen alpha1-CB8. Amino acid
           sequence of the hydroxylamine-produced fragment HA2.
           #cross-references MUID:73006942
           #contents  CNBR8
           #accession A90362
           #molecule_type protein
           #residues  239-418 #label B02
           #experimental_source skin
REFERENCE   #authors  Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
           #journal    Biochemistry (1974) 13:2946-2953
           #title      Chemical studies on the cyanogen bromide peptides of rat skin
           collagen. Amino acid sequence of alpha1-CB3.
           #cross-references MUID:74271984
           #contents  CNBR3
           #accession A90379
           #molecule_type protein
           #residues  419-567 #label B03
           #experimental_source skin
REFERENCE   #authors  Stoltz, M.; Timpl, R.; Furtmayr, H.; Kuehn, K.
           #journal    Eur. J. Biochem. (1973) 37:287-294
           #title      Structural and immunogenic properties of a major antigenic
           determinant in neutral salt-extracted rat-skin collagen.
           #cross-references MUID:74011954
           #contents  CNBR6
           #accession A91209
           #molecule_type protein
           #residues  568-651 #label S01
           #experimental_source skin
           #note       this region probably corresponds to positions 949-1032
           of the alpha 1(I) chain
           the major antigenic determinant (of neutral
           salt-extracted rat skin collagen) involves at least
           Tyr-Asp at positions 650-651 above and probably the
           nine residues that precede the dipeptide
REFERENCE   #authors  Stoltz, M.; Timpl, R.; Kuehn, K.
           #journal    FEBS Lett. (1972) 26:61-65
           #title      Non-helical regions in rat collagen alpha1-chain.
           #cross-references MUID:73049495
           #contents  CNBR6
           #accession A91385
           #molecule_type protein
           #residues  651-671 #label S02
           #experimental_source skin
           #note       the composition of peptides comprising residues 1-9 and
           1-19 confirms the sequence
           this region (residues 651-671 above) probably

```

corresponds to positions 1032-1052 of the alpha 1(I) chain

COMMENT Prolines and lysines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated to varying extents. Prolines are predominantly 4-hydroxylated. Lysines are 5-hydroxylated and subsequently O-glycosylated.

COMMENT The order of the nine CNBR peptides in the alpha 1(I) chain of rat skin collagen was determined as 0,1,2,4,5,8,3,7,6.

COMMENT The complete chain contains 1052 residues.

CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; von Willebrand factor type C repeat homology

KEYWORDS coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

FEATURE 1

9 #modified_site blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status experimental\

103,424,547 #modified_site allysine (Lys) #status experimental\

103 #binding_site carboxylate (Lys) (covalent) #status experimental\

424,547 #modified_site 5-hydroxylysine (Lys) #status experimental\

SUMMARY #modified_site 5-hydroxylysine (Lys) #status experimental\

#length 671 #checksum 936

Insufficient memory To Reconstruct alignment.

RESULT 9 PC4356 #type fragment

ENTRY

Note: remainder of annotations omitted.

Query Match 3.98; Score 180; DB 2; Length 99;

Best Local Similarity 39.28; Pred. No. 5.24e-05;

Matches 51; Conservative 22; Mismatches 57; Indels 0; Gaps 0;

UD E D G G Q D G D Q V G H E D L Y R Y G

Dt 40 GARGAVGNGNGNARGAVGNGNARGAVGNGNARGAVGNGNARGAVGNGN 99

QY 310 GAGCTGTCCAGATGAGCGCCGAGAGTGTGAGCGAGAGTGTGAGCG 369

Qt E P V A D E A D E V E Q K Q L Q K R F G

Db G F L R R I R P K L K W D N O K R Y G G

Dt 100 GGTATTNN 159

QY 370 GGCTTCTGCGGCGCCGAGAGTGTGAGCGAGAGTGTGAGCG 429

Qt G F T G A R K S A R K L A N Q K R F S E

Db F L R

Dt 160 TTYTNNNGM 169

QY 430 TTTATGAGC 439

Qt F M R

RESULT 8

ENTRY

CGRTIS #type fragments

collagen alpha 1(I) chain - rat (fragments)

ORGANISM #formal_name Rattus norvegicus #common_name Norway rat

DATE 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_Change 19-Oct-1995

ACCESSIONS A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91365; A02854; A02855

REFERENCE A90552

#authors Bornstein, P.

#journal Biochemistry (1967) 8:63-71

#title Comparative sequence studies of rat skin and tendon collagen. II. The absence of a short sequence at the amino terminus of the skin alpha1 chain.

#cross-references M01D:69155173

#contents CNBR0 and CNBR1

#accession A90559

#molecule_type protein

#residues 1-19 #label B01

##experimental_source tendon

##note sequences from skin and tendon appear to be identical the amino-terminal tetrapeptide may be removed by limited proteolysis during extraction

REFERENCE A90552

#authors Kang, A.H.; Bornstein, P.; Piez, K.A.

#journal Biochemistry (1967) 6:788-795

#title The amino acid sequence of peptides from the cross-linking region of rat skin collagen.

#cross-references M01D:67162268

#contents CNBR1

#accession A90552

#molecule_type protein

#residues 5-19 #label KAN

##experimental_source skin

REFERENCE A92029

#authors Bornstein, P.

#journal J. Biol. Chem. (1967) 242:2572-2574

#title The incomplete hydroxylation of individual prolyl residues in collagen.

#cross-references M01D:67165368

#contents CNBR2

#accession A92029

#molecule_type protein

#residues 20-55 #label B02

##experimental_source skin and tendon

REFERENCE A90353

#authors Butler, W.T.; Ponds, S.L.

#journal Biochemistry (1971) 10:2076-2081

#title Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of alpha1-CB4.

#cross-references M01D:71263178

#contents CNBR4

#accession A90353

#molecule_type protein

#residues 56-102 #label B01

##experimental_source skin

REFERENCE A90566

#authors Butler, W.T.

#journal Biochemistry (1970) 9:44-50

#title Chemical studies on the cyanogen bromide peptides of rat skin collagen. The covalent structure of alpha1-CB5, the major hexose-containing cyanogen bromide peptide of alpha1.

#cross-references M01D:70085124

#contents CNBR5

#accession A90566

#molecule_type protein

#residues 103-139 #label B02

##experimental_source skin

REFERENCE A90357

#authors Ballan, G.; Click, E.M.; Bornstein, P.

#journal Biochemistry (1971) 10:4470-4478

#title Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylamine-produced fragment HA1.

#cross-references M01D:72136131

#contents CNBR8

#accession A90357

#molecule_type protein

#residues 140-238 #label BA1

##experimental_source skin

REFERENCE A90362

#authors Ballan, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.

#journal Biochemistry (1972) 11:3798-3806

#title Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylamine-produced fragment HA2.

#cross-references M01D:73006942

#contents CNBR8

#accession A90362

#molecule_type protein

#residues 239-418 #label BA2

```

REFERENCE      #experimental_source skin
#authors      Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
#journal      Biochemistry (1974) 13:2946-2953
#title        Chemical studies on the cyanogen bromide peptides of rat skin
               collagen. Amino acid sequence of alpha1-CB3.
#cross-references MUID:74271964
#contents     CNB73
#accession    A90379
#molecule_type protein
#residues     419-567 #label BU3
#experimental_source skin

REFERENCE      #experimental_source skin
#authors      Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
#journal      Eur. J. Biochem. (1973) 37:287-294
#title        Structural and immunogenic properties of a major antigenic
               determinant in neutral salt-extracted rat-skin collagen.
#cross-references MUID:74011954
#contents     CNB76
#accession    A91209
#molecule_type protein
#residues     568-651 #label SM1
#experimental_source skin
#note         this region probably corresponds to positions 949-1032
               of the alpha 1(I) chain
               the major antigenic determinant (of neutral
               salt-extracted rat skin collagen) involves at least
               Tyr-Asp at positions 650-651 above and probably the
               nine residues that precede the dipptide

REFERENCE      A91385
#authors      Stoltz, M.; Timpl, R.; Kuehn, K.
#journal      FEBS Lett. (1972) 26:61-65
#title        Non-helical regions in rat collagen alpha1-chain.
#cross-references MUID:73049495
#contents     CNB76
#accession    A91385
#molecule_type protein
#residues     651-671 #label SM2
#experimental_source skin
#note         the composition of peptides comprising residues 1-9 and
               1-19 confirms the sequence
               this region (residues 651-671 above) probably
               corresponds to positions 1032-1052 of the alpha 1(I)
               chain

COMMENT        Prolines and lysines at the third position of the tripeptide
               repeating unit (G-X-Y) are hydroxylated to varying extents.
               Prolines are predominantly 4-hydroxylated. Lysines are
               5-hydroxylated and subsequently O-glycosylated.
               The order of the nine CNB7 peptides in the alpha 1(I) chain of rat
               skin collagen was determined as 0,1,2,4,5,8,3,7,6.
COMMENT        The complete chain contains 1052 residues.
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
               carboxy-terminal homology; von Willbrand factor type C
               repeat homology
KEYWORDS        coll; extracellular matrix; glycoprotein; trimer;
               triple helix

FEATURE        #modified_site blocked amino end (Glx) (probably
1               pyroglutamate carboxylic acid) #status experimental\
9               #modified_site allysine (Lys) #status experimental\
103,424,547     #binding_site carboxylate (Lys) (covalent) #status
               experimental\
103             #modified_site 5-hydroxylysine (Lys) #status
               experimental\
424,547         #modified_site 5-hydroxylysine (Lys) (partial) #status
SUMMARY        #length 671 #checksum 936
               Insufficient memory To Reconstruct alignment.

```

```

TITLE          prodynorphin - mouse (fragment)
ORGANISM       #formal_name Mus musculus #common_name house mouse
DATE           02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
                26-Feb-1998
ACCESSIONS     PC4356; PC4357
REFERENCE      #authors      Gulya, K.; Orpena, A.K.; Sikelä, J.M.; Hoffman, P.L.
                #journal      Mol. Brain Res. (1993) 20:1-8
                #title         Prodynorphin and vasopressin mRNA levels are differentially
                                affected by chronic ethanol ingestion in the mouse.
#accession     PC4356
#molecule_type DNA
#residues     1-63 ##label GUL
#accession     PC4357
#molecule_type protein
#residues     1-7:57-63 ##label GUT
#experimental_source brain
CLASSIFICATION #superfamily proenkephalin
FEATURE
1-4            #product alpha-neoendorphin #status predicted #label
                ANE\
32-36,51-55   #product leucine endorphalin #status predicted #label
                LEE\
37-48          #product dynorphin-A #status predicted #label DNA\
56-63          #product dynorphin-B #status predicted #label DNB
                #length 63 #checksum 8609
SUMMARY
Query Match    3.9%; Score 180; DB 2; Length 771;
Best Local Similarity 39.2%; Pred. No. 5 24e-05;
Matches        51; Conservative 22; Mismatches 57; Indels 0; Gaps 0

D  E D G G Q D G D Q V G H E D L Y K R Y G
D  40 GARGAYGNGGNCARGNAYCARGTGNCACAYGAYTNTAYARBGNTAYGNG 99
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
O  310 GAGCCTGTCGACGAGATGACGCGATGAGGAGGACGAGACCTCAGAAAGTTGGG 369
   E P V A D E A D E V E Q K Q L Q K R F G

D  G F L R R I R P K L K W D N Q K R Y G G
D  100 GNTTYTYNMGNGMAGTTHMGNCNARNTARNTAGTGAGVAAACARAAARMGNTAYGNGNG 159
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
O  370 GGCTTCACTGGGGCCGAGACGACCGGAGAGTGGCCACACAGAGCGTTCAGTGAG 429
   G F T G A R K S A R K L A N Q K R F S E

D  F L R
D  160 TTYTYNMGNM 169
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
O  430 TTTATGAGGC 439
   F M R

RESULT 9
ENTRY   PC4356 #type fragment
TITLE   prodynorphin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
         26-Feb-1998
ACCESSIONS PC4356; PC4357
REFERENCE  #authors      Gulya, K.; Orpena, A.K.; Sikelä, J.M.; Hoffman, P.L.
           #journal      Mol. Brain Res. (1993) 20:1-8
           #title         Prodynorphin and vasopressin mRNA levels are differentially
                           affected by chronic ethanol ingestion in the mouse.
#accession PC4356
#molecule_type DNA
#residues 1-63 ##label GUL
#accession PC4357
#molecule_type protein
#residues 1-7:57-63 ##label GUT
#experimental_source brain
CLASSIFICATION #superfamily proenkephalin
FEATURE
1-4            #product alpha-neoendorphin #status predicted #label
                ANE\

```



```

OY      215 ACCAGTCGAAAGCCTCGCATATCAGACACCTGAGAGAACCCGCGTGTAGAGATGTGG 274
        P V E S L G D A A P E E N A A C Q E V V
Dt      812 ARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARG 871
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      275 TGCAGCCCGACGACCGACCTGAGGACATGACAGCGCTGCGCATGAGAGCCGATG 334
        Q A R D A E P E A D A E P V A D E A D E
Db      872 AYGARGNGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARGARG 888
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      335 AGCTGAGCAGAACGAC 351
        V E Q K Q

RESULT  12
ENTRY   VIGH2      #type complete
TITLE   vitellogenin II precursor - chicken
CONTAINS #protein: 1160vitellin; phospholip
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE     28-Aug-1985 #sequence_revision 30-Jun-1988 #text_change
        20-Mar-1998
ACCESSIONS
A92941: A93502; C29184; I50440; S07788; A03335; A23177;
A27869
REFERENCE
#authors van het Schip, F.D.; Samalio, J.; Broos, J.; Ophuis, J.;
#journal J. Mol. Biol. (1987) 196:245-260
#title   Nucleotide sequence of a chicken vitellogenin gene and
        derived amino acid sequence of the encoded yolk precursor
        protein.
#cross-references M01D:88011328
#accession A92941
#molecule-type DNA
#residues 1-1650 ##label VAN
#cross-references GB:X13607; NID:g63886; PID:g63887
#note    1840-Ale was also found
REFERENCE
A93502
#authors Burch, J.B.E.
#journal Nucleic Acids Res. (1984) 12:1117-1135
#title   Identification and sequence analysis of the 5' end of the
        major chicken vitellogenin gene.
#cross-references M01D:84118805
#accession A93502
#molecule-type DNA
#residues 1-71 ##label BUR
#cross-references GB:X00345; NID:g63872; PID:g63873
REFERENCE
A91754
#authors Clark, R.C.
#journal Int. J. Biochem. (1985) 17:983-988
#title   The primary structure of avian phosphatins. Contributions
        through the Edman degradation of methylmercaptatins
        prepared from the constituent phosphoproteins.
#cross-references M01D:86056531
#accession C29184
#molecule-type protein
#residues 1112-1188 ##label CLA
REFERENCE
I50440
#authors Byrne, B.M.; van Het Schip, A.D.; van de Klundert, J.A.
#journal Biochemistry (1984) 23:4275-4279
#title   Amino acid sequence of phosphatins derived from the nucleotide
        sequence of part of the chicken vitellogenin gene.
#cross-references M01D:85023316
#accession I50440
#status translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 1092-1338 ##label BYR
#cross-references GB:K02113; NID:g212878; PID:g212879
REFERENCE
S07788
#authors Philippaen, J.N.J.; de Vries, J.E.; Samalio, J.; van Dijk, C.;
        Arnborg, A.C.; AB, G.

```

```

#journal J. Mol. Evol. (1989) 28:185-190
#title Characterization of a polymorphism in the 3' part of the
        chicken vitellogenin gene.
#cross-references M01D:89178745
#accession S07788
#status translation not shown
#molecule-type DNA
#residues 1705-1757 ##label ABC
#cross-references EMBL:X14729; NID:g63881; PID:e13369; PID:g1334751
#note    The source is designated as Gallus domesticus
COMMENT   Vitellogenin is synthesized in the liver of oviparous vertebrates
        in response to steroid (estrogen) induction.
COMMENT   Phosphatins, an egg yolk storage protein, is one of the most highly
        phosphorylated (10%) proteins in nature.
GENETICS
#gene
#introns 14/1: 21/1; 71/3: 155/3; 209/3: 309/3; 362/2: 417/3;
        459/3; 532/3; 605/2; 638/1; 679/3; 713/3; 753/3; 817/1;
        848/2; 910/2; 958/3; 1033/1; 1091/3; 1321/3; 1338/5;
        1368/3; 1439/3; 1487/2; 1533/3; 1579/1; 1641/2; 1672/1;
        1704/3; 1758/1; 1812/1
CLASSIFICATION #superfamily vitellogenin
KEYWORDS egg yolk; glycoprotein; liver; phosphoprotein
FEATURE
1-15
1112-1328
1123,1124,1125,
1126,1127,1128,
1130,1131,1135,
1136,1137,1138,
1139,1140,1142,
1143,1167,1168,
1169,1170,1172,
1173,1175,1176,
1178,1179,1182,
1183,1184,1186,
1187
1189,1190,1193,
1194,1195,1196,
1197,1198,1199,
1200,1201,1202,
1203,1204,1206,
1207,1208,1209,
1210,1211,1212,
1213,1214,1215,
1217,1219,1220,
1221,1222,1223,
1224,1226,1227,
1228,1229,1230,
1231,1233,1235,
1236,1237,1239,
1240,1241,1242,
1243,1244,1245,
1246,1247,1248,
1249,1250,1251,
1252,1254,1255,
1256,1257,1259,
1260,1261,1262,
1263,1264,1266,
1267,1268,1271,
1273,1276,1282,
1283,1284,1285,
1286,1287,1288,
1289,1291,1293,
1296
1280
1301,1308,1309,
1310,1311,1312,
1313,1314,1315
#binding_site phosphate (Ser) (covalent) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
predicted\
#binding_site carbohydrate (Asn) (covalent) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
predicted

```



```

RESULT 15
ENTRY
TITLE DFPC #type complete
ALTERNATE_NAMES beta-neoendorphin / dynorphin precursor - pig
CONTAINS prenekephalin B precursor
          alpha-neoendorphin; beta-neoendorphin; dynorphin;
          leu-enkephalin; leumorphin; rimorphin
          #formal_name Sus scrofa domestica #common_name domestic pig
          15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change
          19-Apr-1996
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:82220112
#accession A93280
#molecule_type mRNA
#residues 1-256 #label NAK
#experimental_source hypothalamus
#note five nucleotide differences were found in the protein
      coding region among three of the clones used. Two of
      these result in sequence differences: Gly or Glu at
      position 147 and Arg or Gly at position 197
REFERENCE
#authors
#journal
#title
#cross-references MUID:84127982
#accession A90117
#molecule_type protein
#residues 228-256 #label NAK
REFERENCE
#authors
#journal
#title
#cross-references MUID:83039385
#accession A19145
#molecule_type protein
#residues 209-240 #label FIS
COMMENT Leumorphin acts as an agonist at the kappa opioid receptor.
CLASSIFICATION #superfamily proenkephalin
KEYWORDS hypothalamus; neuropeptide; opioid peptide; polymorphism
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
175-183 #product alpha-neoendorphin #status predicted #label
209-240 #product beta-neoendorphin #status predicted #label BNF\
209-225 #product dynorphin, long form #status experimental
228-256 #product dynorphin A #status experimental #label DNF\
228-240 #product leumorphin #status experimental #label IMF\
228-232 #product rimorphin (dynorphin B) #status experimental
          #label RMF\
SUMMARY #product leu-enkephalin #status experimental #label LEN
          #length 256 #molecular_weight 28616 #checksum 8694
QUERY Match 3.6% Score 170; DB 1; Length 251;
Best Local Similarity 39.4%; Pred. No. 8,75e-04;
Matches 43; Conservative 20; Mismatches 46; Indels 0; Gaps 0;
DB D K V G H E D L Y K R Y G G F L R R I R
592 GAYARGTNGNCAYGAGAYTNTAYARAGNTAYGNGNTYTNMGNMGNTHGNGN 651
||:||||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 331 GATGAGTGGAGACAGACACTGCGAAGAGTGTGGCGCTTACACGCGCGGAG 390
Qc D E V E Q K Q L Q K R F G G F T G A R K

```

```

DB P K L K W D N Q K R Y G G F L R
DC 652 CQNAFTNARTGGAYAYCARPARAGNTANGNGNTTYTNMGNM 700
||:||||| ||:| ||:| ||:| ||:| ||:| ||:|
QY 391 TCAGCCGGAAGTGTGGCGACAGACAGCGTTCGAGTATGAGGC 439
Qc S A R K L A N Q K R F S E F M R
RESULT 15
ENTRY
TITLE HBVOY #type complete
ALTERNATE_NAMES hemoglobin beta chain - yellow-cheeked vole
CONTAINS #formal_name Microtus xanthognathus #common_name
          yellow-cheeked vole
          29-Jul-1981 #sequence_revision 29-Jul-1981 #text_change
          14-Nov-1997
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#accession A02408
#molecule_type protein
#residues 1-146 #label DUF
CLASSIFICATION #superfamily globin; globin homology
KEYWORDS blood; chromoprotein; erythrocyte; heme; iron; oxygen carrier
FEATURE
3-146 #domain globin homology #label GLB\
63 #binding_site oxygen (His) (distal axial ligand) #status
92 #binding_site heme iron (His) (proximal axial ligand)
SUMMARY #length 146 #molecular_weight 15672 #checksum 184
Insufficient memory To Reconstruct alignment.
RESULT 20
ENTRY
TITLE DFHU #type complete
ALTERNATE_NAMES beta-neoendorphin / dynorphin precursor - human
CONTAINS prenekephalin B precursor
          beta-neoendorphin; dynorphin; leu-enkephalin; leumorphin;
          rimorphin
          #formal_name Homo sapiens #common_name man
          03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change
          20-Mar-1998
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:84068211
#accession A01478
#molecule_type DNA
#residues 1-254 #label HOR
GENETICS #cross-references GB:K02268; GB:X00177; NID:g182103; PID:g182105
          #gene GDB:PDYN
          #map_position 20pter-20p12
          #introns 44/3
          #note the intron shown is the third intron
CLASSIFICATION #superfamily proenkephalin
KEYWORDS neuropeptide
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
175-183 #product beta-neoendorphin #status predicted #label BNF\
209-223 #product dynorphin #status predicted #label DNF\
226-254 #product leumorphin #status predicted #label IMF\
226-238 #product rimorphin #status predicted #label RMF\
226-230 #product leu-enkephalin #status predicted #label LEN

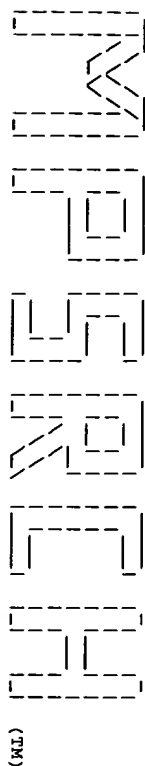
```

```
SUMMARY
#length 254  #molecular-weight 28385  #checksum 5710
```

Query Match	3.5%;	Score 165;	DB 1;	Length 213;
Best Local Similarity	38.9%;	Pred. No. 3.43e-03;		
Matches	49;	Conservative	20;	Mismatches 57;
			Indels	0;
			Gaps	0;

Dd	G E E G D G D S M G H E D L Y K R Y G G F	
Dt	569 CNGSGANGGNGAYGNGNYSMTGGNGCAGRGAYTTTATAAAGNGTAIGNGNT	6288
Oy	314 CTGTGCGCATGAGCCGATAGGTGGAGCAGAAAGCAGCTGACAAAAAGGTGGGGCT	3737
Qt	V A D E A D E V E Q K Q L Q K R F G G G F	
Dd	L R R I R P K L K W D N O K R Y G G F L	
Dt	629 TTYTMMGNMNAHTHNGCNMAARTNARTEGGAYAAACARAPMGTATAGNGNGNTTY	6688
Oy	374 TCACGGGGGGCCGAGTCACCCGGAAGGTGGCGCAACCAACCGGTTCAGTAGTTTA	4333
Qt	T G A R K S A R K L A N Q K R F S E F M	
Dd	R	
Dt	689 TTMGNM 694	
Oy	1:1:	
Qt	434 TGAGGC 439	
	R	

Search completed: Fri Apr 16 14:41:23 1999
Job time : 420 secs.



(TM)

Release 3.1a John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Run on: Tue Apr 27 11:04:21 1999; Maspar time 1255.89 Seconds
1328.940 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-1
Description: (1-932) from US09011797.seq
Perfect Score: 932
N.A. Sequence: 1 CTCCTCCAGCGCTTCAGCAG.....AAAAAAGGAATTC 932
Comp: GAGAGGTCGCACAGTCGTC.....TTTTTTTTTTCCTTAAG

Scoring table: TABLE default

Gap 6

Match STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

emb1-est56

Database:

emb1-est56

1:em-est1 2:em-gss1 3:em-gss2 4:em-gss3
genbank-est109

5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13
10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17
14:gb-est18 15:gb-est19 16:gb-est2 17:gb-est20
18:gb-est21 19:gb-est3 20:gb-est4 21:gb-est5 22:gb-est6
23:gb-est7 24:gb-est8 25:gb-est9 26:gb-gss1 27:gb-gss2
28:gb-gss3 29:gb-gss4

Statistics: Mean 11.547; Variance 3.810; scale 3.031

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
C 1	377	40.5	382	15	AI044049	UI-R-C1-jv-f-08-0-UI.s	0.00e+00
C 2	355	38.1	621	18	AI178977	EST1222659 Normalized r	0.00e+00
C 3	311	33.4	374	14	AA801447	EST190944 Normalized r	0.00e+00
C 4	221	23.7	362	20	W50611	md04d04.r1 Soares mous	2.49e-227
C 5	193	20.7	372	18	AI153441	uc35c08.r1 Soares 2NBH	1.79e-192
C 6	176	18.9	249	22	AA023304	mh70d02.r1 Soares mous	1.87e-171
C 7	166	17.8	445	10	AA648997	ns41g11.s1 NCI_CGAP_GC	3.73e-159
C 8	162	17.4	319	13	AA845221	ak76d12.s1 Bartshead sp	3.00e-154
C 9	155	16.6	365	13	AA760677	n209c12.s1 NCI_CGAP_GC	1.10e-145
C 10	134	16.5	397	8	AA283020	ztl6b04.s1 NCI_CGAP_GC	1.83e-144
C 11	154	16.5	431	8	AA282757	z891a07.s1 NCI_CGAP_GC	1.83e-144
C 12	154	16.5	431	13	AA806350	oc26f09.s1 NCI_CGAP_GC	1.83e-144
C 13	151	16.2	344	12	AA809444	ob43e01.s1 NCI_CGAP_GC	8.33e-141

C 14	146	15.7	305	7	HUM116D04A	Human fetal brain cDNA	1.01e-134
C 15	132	14.2	499	14	AA801448	EST190945 Normalized r	8.90e-118
C 16	131	14.1	364	16	AA51100	y938b10.s1 Homo sapien	1.42e-116
C 17	117	12.6	392	8	AA282862	z591a08.r1 NCI_CGAP_GC	8.26e-100
C 18	114	12.2	186	20	HSAAACBIR	H. sapiens putatively	3.05e-96
C 19	100	10.7	199	20	W77365	m655b09.r1 Soares mous	9.62e-80
C 20	70	7.5	458	8	AA283106	ztl6b04.r1 NCI_CGAP_GC	1.69e-45
C 21	59	6.3	352	12	AA754459	y938b107 Rice Immature	1.58e-33
C 22	59	6.3	359	12	R19874	y938b107 Rice Immature	1.58e-33
C 23	56	6.0	252	12	AA754459	y938b107 Rice Immature	1.58e-33
C 24	50	5.4	247	12	AA754458	y938b107 Rice Immature	1.58e-33
C 25	49	5.3	247	12	AA754458	y938b107 Rice Immature	1.58e-33
C 26	36	3.9	2275	11	AF034173	Homo sapiens ntcon2 co	1.77e-10
C 27	34	3.6	340	13	AA885285	al58g01.s1 Soares NFL	1.12e-08
C 28	31	3.3	2275	11	AF034173	Homo sapiens ntcon2 co	4.49e-06
C 29	29	3.1	267	14	AA873223	oh70g06.s1 NCI_CGAP_K1	2.07e-04
C 30	29	3.1	356	17	AU015060	Mus musculus 2-cell em	2.07e-04
C 31	29	3.1	356	14	AA835988	msc1908.s1 NCI_CGAP_GC	2.07e-04
C 32	29	3.1	452	8	AA414281	vc60a08.s1 Knowles Sol	2.07e-04
C 33	29	3.1	456	18	AA147376	q964c09.s1 Soares test	2.07e-04
C 34	29	3.1	520	6	AA436394	zy44a03.s1 Soares ovar	2.07e-04
C 35	29	3.1	603	8	AA100987	z414e03.s1 StrataGene	2.07e-04
C 36	28	3.0	223	18	AI136193	UI-R-C2p-rs-d-05-0-UI	1.33e-03
C 37	28	3.0	297	10	AA640712	nr22g05.r1 NCI_CGAP_Pt	1.33e-03
C 38	28	3.0	320	20	W91592	MIA.E03.070.C MTA Adul	1.33e-03
C 39	28	3.0	341	17	AI088019	ov60e10.x1 Soares NSF	1.33e-03
C 40	28	3.0	406	22	AA204983	z571a03.s1 StrataGene	1.33e-03
C 41	28	3.0	433	14	AA830841	oc54c01.s1 NCI_CGAP_GC	1.33e-03
C 42	28	3.0	504	11	AA862291	z185h06.s1 Soares feta	1.33e-03
C 43	28	3.0	505	15	AI036290	ud60f04.x1 Sugano mous	1.33e-03
C 44	27	2.9	477	13	AA837073	oc18d12.s1 NCI_CGAP_GC	8.20e-03
C 45	27	2.9	622	11	HS069183	Human DRES 44 mRNA seq	8.20e-03

ALIGNMENTS

RESULT 1	AI044049	382 bp	MRNA	EST	06-JUL-1998
LOCUS	UI-R-C1-jv-f-08-0-UI.s1	UI-R-C1 Rattus norvegicus	CDNA clone		
DEFINITION	UI-R-C1-jv-f-08-0-UI 3', mRNA sequence.				
ACCESSION	AI044049				
NID	93290952				
KEYWORDS	EST.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
TITLE	1 (bases 1 to 382)				
JOURNAL	Bonaldi,M.F., Lennon,G. and Soares,M.B.				
MEDLINE	Normalization and subtraction: two approaches to facilitate gene				
COMMENT	discovery				
	Genome Res. 6 (9), 791-806 (1996)				
	97044477				

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult brain library. cDNA library Preparation: M. Fatima Bonaldi,
Ph.D. Clone distribution: clones will be available through Research
Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1..382
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/note="Vector: pRT3D-Pac (pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-C1

library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dr track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

/db.xref="taxon:10116"
/clone="UI-R-C1-jv-f-08-0-UI"
/clone.lib="UI-R-C1"
/dev.stage="Adult"
/lab_host="DH10B (Life Technologies)"

BASE COUNT 100 a 105 g 102 t 1 others
ORIGIN

Query Match 40.5%; Score 377; DB 15; Length 382;
Best Local Similarity 99.2%; Pred. No. 0.00e+00;

Matches 379; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 TTTTATTTTATTTTACATTTGCTCAGTAGAAGTTAATGAGAAAGTTGTTCA 60
|||||
Cp 922 TTTTATTTTATTTTACATTTGCTCAGTAGAAGTTAATGAGAAAGTTGTTCA 863
Db 61 AACATCACTCAAAAAGACAGCTCCTTTACAACCAATTATGCGAGTGGAGCAAA 120
|||||
Cp 862 AACATCACTCAAAAAGACAGCTCCTTTACAACCAATTATGCGAGTGGAGCAAA 803
Db 121 CCCAGGTGATTTATTTCTCTCACCCTGCTACGAGAGGCGAGAGATGACAA 180
|||||
Cp 802 CCCAGGTGATTTATTTCTCTCACCCTGCTACGAGAGGCGAGAGATGACAA 743
Db 181 GAAGATGGGCGCATTTAAATGTTAGCTGAGGAATTAACAAAATACACTGATGAGCA 240
|||||
Cp 742 GAAGATGGGCGCATTTAAATGTTAGCTGAGGAATTAACAAAATACACTGATGAGCA 683
Db 241 ACAGATTTGTTGAGACATGCTGTGGGAGAGTGCAGGCTGTTCAAGTCCACAGGAA 300
|||||
Cp 682 ACAGATTTGTTGAGACATGCTGTGGGAGAGTGCAGGCTGTTCAAGTCCACAGGAA 623
Db 301 GGAGGAATCTCTACCTCCGCGCTGCAAGTCTTGTGTGACACATGCTGTACGGGGCTC 360
|||||
Cp 622 GGAGGAATCTCTACCTCCGCGCTGCAAGTCTTGTGTGACACATGCTGTACGGGGCTC 563
Db 361 ACCTGATGCTCATGGGTGCA 382
|||||
Cp 562 ACCTGATGCTCATGGGTGCA 541

RESULT 2
LOCUS A1178977 621 bp mRNA EST 08-OCT-1998
DEFINITION EST222659 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
ACCESSION RSPBT94.3 end, mRNA sequence.
NID A1178977
KEYWORDS 93729615
SOURCE EST.
Rattus sp.

ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 621)
AUTHORS Lee,N.H., Gloder,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlanage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: TC52405
Contact: Lee, NH
ATCC

FEATURES
Source
Location/Qualifiers
1..621
/organism="Rattus sp."
/note="Organ: spleen; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"
/db.xref="taxon:10118"
/db.xref="RSPBT94"
/clone.lib="Normalized rat spleen, Bento Soares"
BASE COUNT 140 a 166 c 166 g 149 t
ORIGIN

Query Match 38.1%; Score 355; DB 18; Length 621;
Best Local Similarity 98.1%; Pred. No. 0.00e+00;

Matches 362; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1 ATTAACATTTTGTCTCAGTAGAAGTTAATGAGAAAGTTGTTCAACATCACTCAAA 60
|||||
Cp 908 ATTAACATTTTGTCTCAGTAGAAGTTAATGAGAAAGTTGTTCAACATCACTCAAA 849
Db 61 AAGAAGAGCTCCTTTACAACCAATTATGCGAGTGGAGCAAAACCCAGGTGATT 120
|||||
Cp 848 AAGAAGAGCTCCTTTACAACCAATTATGCGAGTGGAGCAAAACCCAGGTGATT 789
Db 121 TCATGTCCTCTCACCTGCGCTACGAGAGGCGAGAGATGACGAAGATGGGCGCAT 180
|||||
Cp 788 TCATGTCCTCTCACCTGCGCTACGAGAGGCGAGAGATGACGAAGATGGGCGCAT 729
Db 181 TAAATGTTAGCTGAGGAATTTACAAAATACACTGATGATAGCAACAGATTGCTG 240
|||||
Cp 728 TAAATGTTAGCTGAGGAATTTACAAAATACACTGATGATAGCAACAGATTGCTG 669
Db 241 AGACATGCTGTGGGAGAGTGCAGGCTGTTCAAGTCCACAGGAGGAATCTGAC 300
|||||
Cp 668 AGACATGCTGTGGGAGAGTGCAGGCTGTTCAAGTCCACAGGAGGAATCTGAC 609
Db 301 TCCGCGCTGAGAGCTGTGTGTGAGACATGCTGTACGGGGCTCACCTGGATGCAATT 360
|||||
Cp 608 TCCGCGCTGAGAGCTGTGTGTGAGACATGCTGTACGGGGCTCACCTGGATGCAATT 549
Db 361 CTGCTGACAG 369
|||||
Cp 548 GGGTTCAG 540

RESULT 3
LOCUS AA801447 374 bp mRNA EST 30-APR-1998
DEFINITION ESP190944 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
ACCESSION AA801447
NID 92864402
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 374)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat
JOURNAL Unpublished (1998)
COMMENT Other ESTs: EST190945
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igrr.org
Seq primer: M13-21.
Location/Qualifiers
1. 374
/organism="Rattus sp."
/note="Organ: spleen; Vector: pT7T3pac; Site_1: EcoRI;
Site_2: NotI"
/db_xref="ATCC (inhost):2009517"
/db_xref="taxon:10118"
/clone="RSPAA76"
/clone.lib="Normalized rat spleen, Bento Soares"
BASE COUNT 101 a 82 c 107 g 84 t
ORIGIN
Query Match 33.4%; Score 311; DB 14; Length 374;
Best Local Similarity 98.4%; Pred. No. 0.00e+00;
Matches 316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 1 CTAGAAATTATGAGAAAGTTGTTCAACATCACTCAAAAGAAGACCTCTTAC 60
CP 891 GTAAATTTATGAGAAAGTTGTTCAACATCACTCAAAAGAAGACCTCTTAC 832
Db 61 AACACATTATGAGAGTGTGAGCAAAACCCAGTGTATTCATGTTCTCTCACT 120
CP 831 AACACATTATGAGAGTGTGAGCAAAACCCAGTGTATTCATGTTCTCTCACT 772
Db 121 GGGCCTACGAGAGGAGAGAGATGAGCAAGAAGATGGGGCATTAAATGTTAGCTGAG 180
CP 771 GGGCCTACGAGAGGAGAGAGATGAGCAAGAAGATGGGGCATTAAATGTTAGCTGAG 712
Db 181 GAATTACAAAAATCACTGTATGTATGACACAGATTTGTGTGACATGCTGTGGGAG 240
CP 711 GAATTACAAAAATCACTGTATGTATGACACAGATTTGTGTGAGATGCTGTGGGAG 652
Db 241 GTGGCGGGGGTGTCAAGGCTCAGGGAAGAGAGATCTCACTCCGGGCAATGCATGCT 300
CP 651 GTGGCGGGGGTGTCAAGGCTCAGGGAAGAGAGATCTCACTCCGGGCAATGCATGCT 592
Db 301 TGTGTGAGAGATGCGTTAC 321
CP 591 TGTGTGAGAGATGCGTTAC 571
RESULT 4
LOCUS M50641 362 bp mRNA EST 29-MAY-1996
DEFINITION M004004.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
ACCESSION W50641
NID g1339114
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Rodentia; Sciurognathi; Myomorphia; Muridae;
Murinae; Mus.
1 (bases 1 to 362)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Stepien,M., Tan,F., Underwood,K., Moore,B.,
Theisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterson,R.
TITLE The WashU-HM1 Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:228831
Seq primer: mob.REGA+ET
High quality sequence stop: 351.
Location/Qualifiers
1. 362
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGAGTGGAGCGCGCGGGAATTTTATTTTATTTTATTTT
T 3'/], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="367399"
/clone.lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
<1. >362
BASE COUNT 76 a 121 c 89 g 76 t
ORIGIN
Query Match 23.7%; Score 221; DB 20; Length 362;
Best Local Similarity 90.2%; Pred. No. 2.49e-227;
Matches 248; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Db 88 CTCGCCAGCGTGTTCAGACACTGTCCAGGACTGCTCACTGACCGAGAAAGCTCAC 147
QY 1 CTCGCCAGCGTGTTCAGACACTGTCCAGGACTGCTCACTGACCGAGAGAGCTTCAC 60
Db 148 CCAAGCTCCAGACAGCTCAACTTAAGAGTGCATCTCCAGTGTGAGAGAGGCTTC 207
QY 61 CCGGCTCCGGGACAGCTTCAACTGAACTGTGATCTCCAGTGTGAGAGAGGCTTC 120
Db 208 CCCCAGCCCTCTGAGCTGTATGACCAAGATGAGCCAGTGGCTCCGGGACAGCTCAGC 267
QY 121 CCCCAGCCCTCTGAGCTGTATGACCAAGATGAGCCAGTGGCTCCGGGACAGCTCAGC 180
Db 268 CTTGCTACCCAGAGCTTGTGTCACTGCTCTTTACCAAGCAAGGCTCGAGATCAG 327
QY 181 CTTGCTACCCAGAGCTTGTGTCACTGCTCTTTACCAAGCTCGAGATCAG 240
Db 328 CACCTGAAGAGAAATGCCGCTGTCCGAGGCTTGT 362
QY 241 CACCTGAAGAGAAATGCCGCTGTCCGAGGCTTGT 275
RESULT 5
LOCUS A1153441 372 bp mRNA EST 30-SEP-1998
DEFINITION uc5308.r1 Soares 2NDMT Mus musculus cDNA clone 1429358 5', mRNA
ACCESSION A1153441

[illegible]

LOCUS	249 bp	mRNA	EST	21-JAN-1997
DEFINITION	mb70d02.r1 Soares mouse placenta	4NMp13.5	14.5	Mus musculus CDNA
ACCESSION	clone 4456291	5'		mRNA sequence.
AA023504				
NID	91487228			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sclerothalia; Muridae; Murinae;			
REFERENCE	Mus.			
AUTHORS	1 (bases 1 to 249)			
	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theislen,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:273179 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 196. Location/Qualifiers 1. 249 /organism="Mus musculus" /strain="C57BL/6J" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTCACCATCTGATGAGGAGCGCGCCGGAATTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Felima Bonaldo." /db_xref="taxon:10090" /clone="456291" /clone_lib="Soares mouse placenta 4NMp13.5 14.5" /sex="unknown" /dev_stage="adult" /lab_host="DH10B" <1. 249			
FEATURES	source			
BASE COUNT	50 a 72 c 78 g 49 t			
ORIGIN	mRNA			
Query Match	18.9%; Score 176; DB 22; Length 249;			
Best Local Similarity	87.6%; Pred. No. 1,876-171;			
Matches	205; Conservative 0; Mismatches 29; Indels 0; Gaps 0;			
Db	15 GCATCTCCACTGTGAAGAGAGGCTTCCCGCCCTCTGTGACTGTATGCACCAAG 74			
QY	92 GCATCTCCACTGTGAAGAGAGGCTTCCCGCCCTCTGTGACTGTATGCACCAAG 151			
Db	75 TCATGCGCAATGCTGTGCGGAGCTCAACCTGCTGACACCAAGCTTGTGACGTGCTC 134			
QY	152 CCATGCGCAATGCTGTGCGGAGCTCAACCTGCTGACACCAAGCTTGTGACGTGCTC 211			
Db	135 TTATACCAAGCAAGAGGCTCGAGATGACGACTGTGAAGAGATGCGGCTGCGGACT 194			
QY	212 TTATACCAAGCAAGAGGCTCGAGATGACGACTGTGAAGAGATGCGGCTGCGGACT 271			
Db	195 TGTGCAAGTGCAGATGACAGAGCTGCGGAGATGCTGAGCTTGGCGCAGATG 248			

0Y 272 TGTGCAAGCCGAGACGAGCTGTGAGCATGCGACAGCTGTGCGACATG 325

RESULT 7
LOCUS AA648997 445 bp mRNA EST 13-NOV-1997
DEFINITION ns41g11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1186244,
mRNA sequence.
ACCESSION AA648997
NID g2575426
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 445)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Insert length: 659 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 419.
Location/Qualifiers
1. 445
/organism="Homo sapiens"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20⁺, 19D-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTGAAGTGGAGCGCCGCTCTATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Benito Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_image="1186244"
/clone_lib="NCI_CGAP_GCB1"
/issue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 130 a 82 c 130 g 102 t 1 others
ORIGIN

Query Match 17.8%; Score 166; DB 10; Length 445;
Best Local Similarity 86.5%; Pred. No. 3.73e-159;
Matches 221; Conservative 0; Mismatches 30; Indels 6; Gaps 5;

D 1 TTTTATTTTACCATTT-GCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 59
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 915 TTTTATTTTACCATTTGCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAA 858
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 60 TCAGTCAAAAAGACAGCTCTTTTACAAACAGTTATGGAGTGGCAGTCAAAAACCCCA 119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 857 TCAATCAAAAAGACAGCTCTTTTACAAACAGTTATGGAGTGGCAGTCAAAAACCCCA 798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 120 GGTTCATATTCCTTATTCCTTACCTGCGCTAGAACAGAGNAGCGCGGTGAGCAGAG 179
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 797 GGTTCATATTCCTTATTCCTTACCTGCGCTAGAACAGAGNAGCGCGGTGAGCAGAG 741
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 180 AGATGGGCTATTGAAATGTAAGTACAGAAATTAACAAATACACTGTGATGACAAAC 239
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 740 AGATGGGCTATTGAAATGTAAGTACAGAAATTAACAAATACACTGTGATGACAAAC 681
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 240 AGGTTGTGTGAAACATGCGAGGGG 266
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

CP 680 AGCATTTGTGTGAGACATGCTGTGGG 654

RESULT 8
LOCUS AA845221 319 bp mRNA EST 04-MAR-1998
DEFINITION AK76612.s1 Barstead spleen HPLRB2 Homo sapiens cDNA clone
IMAGE:1413815 3', mRNA sequence.
ACCESSION AA845221
NID g2931672
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 319)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE Washu-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson R.
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 189.
Location/Qualifiers
1. 319
/organism="Homo sapiens"
/note="Organ: spleen; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker. Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTTCTTTT-
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATGGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified p7T3 vector.
Library constructed by Bob Barstead."
/db_xref="taxon:9606"
/clone_image="1413815"
/clone_lib="Barstead spleen HPLRB2"
/sex="male"
/dev_stage="adult, 17 years"
/lab_host="DH10B"
BASE COUNT 102 a 56 c 87 g 74 t
ORIGIN

Query Match 17.4%; Score 162; DB 13; Length 319;
Best Local Similarity 86.1%; Pred. No. 3.00e-154;
Matches 229; Conservative 0; Mismatches 31; Indels 6; Gaps 5;

D 1 TTTTATTTTACCATTT-GCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAAAT 59
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 914 TTTTATTTTACCATTTTGTCTCAGTAGAAGTTTATGAGAAATCGTTTAAACAAAT 857
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 60 CAGTCAAAAAGACAGCTCTTTTACAAACAGTTATGGAGTGGCAGTCAAAAACCCCA 119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 856 CAGTCAAAAAGACAGCTCTTTTACAAACAGTTATGGAGTGGCAGTCAAAAACCCCA 797
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 120 GTTCAATTCCTTATTCCTTACCTGCGCTAGAACAGAGNAGCGCGGTGAGCAGAG 179
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 796 GTTCAATTCCTTATTCCTTACCTGCGCTAGAACAGAGNAGCGCGGTGAGCAGAG 740
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 180 GATGGGCTATTGAAATGTAAGTACAGAAATTAACAAATACACTGTGATGACAAAC 239
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
C 739 GATGGGCTATTGAAATGTAAGTACAGAAATTAACAAATACACTGTGATGACAAAC 680
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
D 240 GGGTTGTGTGAAACATGCGAGGGG 265
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

CP 679 GGATTGTGGTGAGACATGCTGTGGGG 654

LOCUS	DEFINITION	RESULT
AA760677	365 bp	18-FEB-1998
nr09c12.s1	NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1287286	
mrna sequence.		

ACCESSION	AA/80617
NID	92809607
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE AUTHORS TITLE	JOURNAL COMMENT
Homo sapiens Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 365) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap , National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution; NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/MLND at: www-bio.linnl.gov/dbp/Image/Image.html	

```

Insert Length: 992      Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 363.
      location/Qualifiers
1. .365
FEATURES
      source

```

REFERENCE 1 (bases 1 to 397)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)

 Contact: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 This clone is available royalty-free through LINT ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 685 Std Error: 0.00
 Seq primer: -41ml3 fwd. EF from Amersham.

FEATURES Location/Qualifiers
SOURCE 1.397

FEATURES	Location/Qualifiers
source	1. .397

Cp	841	GCTCCTTTACAAACCAGTATATGCGAGTGGTGAGGGCAAAAACCCAGGTCGTGATTTCATGTT	782
Db	126	CCTTTCACCTGCCCTCTAGAAGCCGAAGAGCGCGGTGCAGCAGGAGCATGGGCTTTGAA	185
Cp	791	CTCTCACCTGCGCCCTACGAGAGGGCC-AGA-GGATGAGCACA-A-GATGGGGCCATTAA	725
Db	186	ATGAGAGCTAGAGCAATTACAAAAATPACACTGTATGTAGCAACAGGGTTCGTGGAAC	245
Cp	724	ATGTTAGCTGAGAGCAATTACAAAAATPACACTGTATGTAGCAACAGGATTTGTGTAGAC	665
Db	246	ATGCCAGGGGG	256
Cp	664	ATGCTGTGGGG	654

RESULT	11		
LOCUS	AA282757	431 bp	mRNA
DEFINITION	Z891807.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:104820 3'	EST	13-AUG-1997
ACCESSION	AA282757		
NID	g1925691		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;		
	Vertebrate; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;		
	Homo.		
REFERENCE	1 (bases 1 to 431)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
JOURNAL	Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1027 Std Error: 0.00 Seq primer: -41m13 fwd. ET from Amersham High quality sequence stop: 400. Location/Qualifiers 1. 431 /organism="Homo sapiens" /note="Vector: pYT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germlinal center B cells by flow sorting (CD20+ IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGTACCAATCTGAGTAGGAGCGCGCCCTCATTTCTTTTTTTTTT- 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo." /db_xref="taxon:9606" /clone="IMAGE:704820" /clone_lib="NCI CGAP GCB1" /tissue_type="germlinal center B cell" /lab_host="DH10B" complement(<1. >431) /db_xref="GDB:5854544"		
BASE COUNT	122 a	85 c	131 g
ORIGIN			93 t

Query Match	16.5%;	Score 154;	DB 8;	Length 431;
Best Local Similarity	85.7%;	Pred. No. 1,83e-144;		
Matches	215;	Conservative 0;	Mismatches 31;	Indels 5;
			Gaps 4;	

6 TTGTGCTAGTAAGTTAATGAGAAACGCTGTTTAAAGAAATCAGCAAAAAGAAC 65

|||||

Cp	899	TTCGCTAGTAGAAGTTTAATGGAGAAN--GTGTGTTTAAACCAATCATCATAAAGAAC	842
-Db	66	GCTCTTTTACAACAAGTATTAGCAGTGCGAAGTCACAAAACCCAGGTCAATTTCTCAT	125
Cp	841	GCTCCTTTACAACAAGTATTAGCAGTGCGAAGTCACAAAACCCAGGTCAATTTCTCAT	782
-Db	126	CCTTTACACTGCCCTCTAGAAAGCGAAGGCGGGGTGCAGCAGAGAGATGGGGCTTTGA	185
Cp	781	CCTTCACACTGCCCTCTAGAAAGCGAAGGCG--AGA--GGATGAGCACAAG--AGATGGGGCCCATTTAA	725
-Db	186	ATGATGATATAGAGAAATATACAAATACTCTGTGTATGACACAGGTTGGTGAAC	245
Cp	724	ATGTTAGCTGGAGCAATTTACAAAAATACACITCTGTATGTATGACACAGGATTTGTGTGAC	665
-Db	246	ATGCCAGGGG	256
Cp	664	ATGCTGTGGG	654
RESULT	12		
LOCUS	AA806350	431 bp	mRNA
DEFINITION	oc6f09.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1350857 3'	EST	25-MAR-1998
ACCESSION	AA806350		
NID	g2875100		
SOURCE	EST.		
ORGANISM	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
	Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	I (bases 1 to 431)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT			
	Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D. CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LZLN at: www-bio.lzlnl.gov/dbfp/image/image.html		
FEATURES			
source			
	Insert Length: 1103 Std Error: 0.00 Seq primer: -40ml3 fwd. EM from Amersham. Location/Qualifiers 1..431 /organism="Homo sapiens" /note="Vector: pRT3D-Pot (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+ IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGTACCACATCTCAAGTGGAGCGCGCCTCATTTTTTTTTTTTTTTT-3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo." /db_xref="taxon:9606" /clone="IMAGE:1350857" /clone_lib="NCI CGAP GCB1" /tissue_type="germinal center B cell" /lab_host="DH10B"		
BASE COUNT	125 a	80 c	133 g
			93 t

	Query Match	15.7%	Score 146;	DB 7;	Length 305;	
	Best Local Similarity 84.2%;	Pred. No. 1.01e-13;				
	Matches 213;	Conservative 4;	Mismatches 29;	Indels 7;	Gaps 3;	
Db	6	TTTGTCAGTAAATTTAATGAGCAAAATCGTTGTTTAAACAAATCAGTCAAAAAGACA	65			
Cp	899	TTTGTCAGTAAATTTAATGAGCAAA--GTTGTTTAAACAAATCAGTCAAAAAGACA	842			
Db	66	GCTCTTTACAAACAGTATGCGAGTGCACAGTCAAAACCCAGGTTCAATTTCTATT	125			
Cp	841	GCTCTTTACAAACAGTATGCGAGTGCACAGTCAAAACCCAGGTTCAATTTCTATT	782			
Db	126	CCTTTCACCCCTCCTAGACAGAGGGGACAGAGCGCGGTGACAGAGAGATGGGCTATTG	185			
Cp	781	CCTTCACCTGGCCCTACAGAGAGGGGACAGAGA---TGAGCAGA-AGATGGGGGCATTA	727			
Db	186	AAATGCTAGCTAGAGGAATTACAAAAATCTCACTCTGABGTAGCAACAGGGCTGTGTGAA	245			
Cp	726	AAATGCTAGCTGAGGAATTACAAAAATCTCACTCTGATGTGTAGCAACAGGATTTGTGTGAG	667			
Db	246	ACATGCCAGGGGG 258				
Cp	666	ACATGCTGTGGGG 654				

Db	363	GTAAAGCATGTGTGCACACCCAGAACCTGCAMCGGGAGTGAAGATTCTCCCTCCCTGA	422
Oy	571	GTAACTAGTGTGTGCACACCCAGAACCTGCACGCCGGAGTGAAGATTCTCCCTCCCTGA	630
Db	423	GGCACTGAACACCGCGGACACTCTCCACAGACTGTCTCACCAATCTGTGTCTACAT	482
Oy	631	GGCACTGAACACCGCGGACACTCTCCACAGACTGTCTCACCAATCTGTGTCTACAT	690
Db	483	CAGAGTGTATTTTGTGA	499
Oy	691	CAGAGTGTATTTTGTGA	707

Search completed: Tue Apr 27 11:26:55 1999
Job time : 1354 secs.

[illegible]

FEATURES	Location/Qualifiers
source	1. .499

```

/organism="Rattus sp."
/notes="Organ: spleen; Vector: p77T3Pac; Site_1: EcoRI
Site_2: NotI"
/db_xref="ATCC (thost):2009517"
/db_xref="taxon:10118"
/clone="RSPA76"
/clone_lb="Normalized rat spleen, Bento Soares"
BASE COUNT      93 a      155 c      165 g      84 t      others
ORIGIN

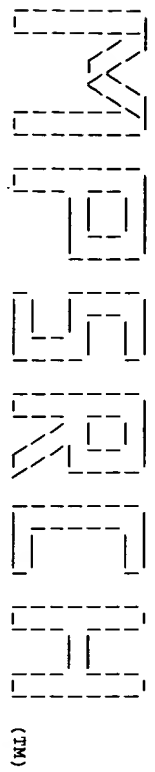
```

Query Match	14.28;	Score 132;	DB 14;	Length 499;
Best Local Similarity	97.88;	Pred. No. 8.90e-118;		
Matches 134;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0

Db	363	GTAAAGCATGTGTGCACACCCAGAACCTGCAMCGGGAGTGAAGATTCTCCCTCCCTGA	422
Oy	571	GTAACTAGTGTGTGCACACCCAGAACCTGCACGCCGGAGTGAAGATTCTCCCTCCCTGA	630
Db	423	GGCACTGAACACCGCGGACACTCTCCACAGACTGTCTCACCAATCTGTGTCTACAT	482
Oy	631	GGCACTGAACACCGCGGACACTCTCCACAGACTGTCTCACCAATCTGTGTCTACAT	690
Db	483	CAGAGTGTATTTTGTGA	499
Oy	691	CAGAGTGTATTTTGTGA	707

Search completed: Tue Apr 27 11:26:55 1999
Job time : 1354 secs.

THIS PAGE BLANK (USPTO)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSearch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:31:00 1999; MasPar time 99.06 Seconds
572.525 Million cell updates/sec

Title: >US-09-011-797-4
Description: (1-8) from US09011797.pep
Perfect Score: 100
N.A. Sequence: 1 ACNNTNCAYCARAYGNAAYGTN 24
Comp: TGNRANGTGTTCRCNTTCAN

Scoring table: TABLE bkttranslated
Gap 40

Mmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_v1
Database: genbank110
16:gb_pat 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl1
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v1

Statistics: Mean 39.423; Variance 98.948; scale 0.398
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	100	100.0	372	29	MMU44027	Mus musculus orphanin
2	100	100.0	633	29	MMPPNMRNA	M.musculus mRNA for pr
3	100	100.0	878	26	HSPPNEX3	H.sapiens gene encodin
4	100	100.0	925	29	S79730	Ortl receptor agonist
5	100	100.0	972	29	RNU48262	Rattus norvegicus pre-
6	100	100.0	988	29	RNPNNRMA	R.norvegicus mRNA for
7	100	100.0	1015	26	HSPNNRMA	H.sapiens mRNA for pre
8	100	100.0	1047	29	MUSN23K	House N23K mRNA for N2
9	100	100.0	1198	28	HSU48263	Human pre-pro-orphanin
10	100	100.0	1253	29	MMPPNEX3	M.musculus gene encodi
11	100	100.0	1354	29	MUSNOP	Mouse mRNA for nocicep
12	95	95.0	1209	20	AB005251	Bos taurus mRNA for no

13	90	90.0	2258	24	SCYBL020W	S.cerevisiae chromosome
14	90	90.0	2551	24	SCU15087	Saccharomyces cerevisi
15	90	90.0	3061	24	YSCHAP	S.cerevisiae HAP3 locu
16	90	90.0	150024	18	HS5222P13	Human DNA sequence ***
17	90	90.0	199092	18	HS80N2	Human DNA sequence ***
18	85	85.0	228	19	EC003705	Elphidium clavatum inc
19	85	85.0	248	29	RRMACPDE05	Rattus norvegicus Wist
20	85	85.0	1080	29	RATPHOCAMB	Rat CAMP phosphodiester
21	85	85.0	1080	29	RATPDE4A	Rat ratPDE4 mRNA encod
22	85	85.0	1086	25	AF071891	Prunus armeniaca 40S r
23	85	85.0	1158	27	AB006077	Homo sapiens doc-1 mRN
24	85	85.0	1193	24	LEGPF46	L.esculentum GPF4 mRNA
25	85	85.0	1704	24	SCU36503	Saccharomyces cerevisi
26	85	85.0	1888	24	SCMR1R30	Saccharomyces cerevisi
27	85	85.0	1888	24	SCMR1RNR	Saccharomyces cerevisi
28	85	85.0	2036	19	AB006675	Halocynthia roretzi Hr
29	85	85.0	2158	22	122474	Sequence 3 from patent
30	85	85.0	2158	29	RATPDP	Rat CAMP phosphodiester
31	85	85.0	2254	21	XIASURAB	Xenopus laevis mRNA fo
32	85	85.0	2263	21	AF019906	Xenopus laevis transme
33	85	85.0	2268	24	SCYDL127W	S.cerevisiae chromosome
34	85	85.0	2647	29	RATPHOSB	Rattus norvegicus phos
35	85	85.0	2814	19	SLMARD	Physarum polycephalum
36	85	85.0	3133	29	RNU95748	Rattus norvegicus CAMP
37	85	85.0	7030	17	SMASSPH	Serratia marcescens DN
38	85	85.0	10510	16	U32761	Haemophilus influenzae
39	85	85.0	12882	16	AE000911	Methanobacterium therm
40	85	85.0	42729	16	MTCY71	Mycobacterium tubercu
41	85	85.0	75317	24	SC4357	S.cerevisiae chromoso
42	85	85.0	103319	27	AC004217	Homo sapiens 12q24.1 p
43	85	85.0	110000	24	SCCHR1	S.cerevisiae chromoso
44	85	85.0	162575	18	AC004086	*** SEQUENCING IN PRO
45	85	85.0	170030	28	AC005829	Homo sapiens Chromosom

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	MMU44027	Mus musculus orphanin	U44027	91335869	house mouse.	Mus musculus	Eukaryote; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 372)	Pan,Y.X., Xu,J. and Pasternak,G.W.	Cloning and expression of a cDNA encoding a mouse brain orphanin	Biochem. J. 315 (Pt 1), 11-13 (1996)	Direct Submission Submitted (28-DEC-1995) Ying-Xian Pan, Neurology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA
								2 (bases 1 to 372)	Pan,Y.-X.			Location/Qualifiers 1..372 /organism="Mus musculus" /db_xref="taxon:10090" /tissue_type="Brain" 2..331 /codon_start=1 /product="orphanin FQ/nociceptin precursor" /db_xref="PID:91335870" /translation="LTSALYQPKASEQHKLMPRVSLVQVRDAEPGADAEFGADAEFGADAEVQKQKRFGEFTGARSKARLANQKRSEFMROYLVLSMQSSQRRRT LHKNNV" misc_feature 188..238

JOURNAL
MEDLINE
REMARK
Nature 377 (6549), 532-535 (1995)
96013745
GenBank staff at the National Library of Medicine created this
entry [NCBI g1dbseq 171655] from the original journal article.

FEATURES This sequence comes from Fig. 4.

source

Location/Qualifiers

1. .925

/organism="Rattus sp."

/db_xref="taxon:10118"

1. .507

/partial

/gene="ORL1 receptor agonist precursor"

1. .507

/partial

/note="endogenous agonist of opioid receptor-like ORL1 receptor. Authors indicate ORL1 receptor agonist peptide is encoded by residues 122-138. Method: conceptual translation with partial peptide sequencing. This sequence comes from Fig. 4"

/product="ORL1 receptor agonist precursor"

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

/translation="LSVFSRSCPEDCLTCOERLHPAPGSFNLKLTICCEKYPFPRPLWICTKAMASDSQSLSPADELTSAALYOSKASEMOKLKPVRSVVQARDEPADAEVDAEADAEVDAEVEKOLKRRGFTGARKSARKLANOKRSEFMROYLYVSMOSSORRRTLHQNGNV"

/codon_start=1

/db_xref="PID:g117239"

3'UTR /product="pro-orphanin FQ"

polya_site

972

/note="45 A nucleotides"

214 a

293 c

251 g

214 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

15; Conservative

5; Mismatches

3; Indels

0; Gaps

0;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

3'UTR /product="pro-orphanin FQ"

polya_site

972

/note="45 A nucleotides"

214 a

293 c

251 g

214 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

15; Conservative

5; Mismatches

3; Indels

0; Gaps

0;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Score 100; DB 29; Length 972;

Pred. No. 5.45e+00;

100.0%;

Best Local Similarity 65.2%; Pred. No. 5,45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 544 ACTGTGACACGAGATGTATGT 566
||:||||:||||:||||:||||:
QY 1 ACNTYCAICARAAAGGNAAYGT 23

RESULT 7
LOCUS HSPBNRMA 1015 bp RNA PRI 10-SEP-1996
DEFINITION H.sapiens mRNA for prepronociceptin.
ACCESSION X97370
NID g1531988
KEYWORDS prepronociceptin.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1015)
Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G.,
Meunier,J.C., and Parmentier,M.
TITLE Structure, tissue distribution, and chromosomal localization of the
prepronociceptin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
MEDLINE 96323281
REFERENCE 2 (bases 1 to 1015)
AUTHORS Parmentier,M.
TITLE Direct Submission
JOURNAL Submitted (19-APR-1996) M. Parmentier, Universite Libre de
Bruxelles, 1 R I B N ULB Campus Erasme, 808 Route de Lennik, B-
1070 Bruxelles, BELGIUM

FEATURES
source location/Qualifiers

1..1015
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="Lambda DASHII (Stratagene #94520)"
/clone="HG91"
/clone="HGEX107"
/chromosome="8"
/map="P21"
/map="1.9 CR from WI-1172 marker"
/number=2
join<1..149,150..601,602..>1015)
24..68
join(24..149,150..554)
/codon_start=1
/product="prepronociceptin"
/db_xref="PID:g244777"
/db_xref="PID:g1531989"
/translation="MKVLICDILLSLVSSVSSCORDCITCQEKIHPALDSFDLEVC
ILCEEKVFPSPILMTCTKVMASWSQSPAPAEHVAALYQPRASEMOHLRMRVR
SLFEQEEPEEPGEAGEEOKOLOKRFGEFTGARKSARKLNOKRFSHFMYLYLS
MOSSORRRLRHONGNV"
69..551
/product="prepronociceptin"
150..601
/number=3
602..>1015
/number=4
1015

mat_peptide

exon

polyA_site 212 a 304 c 264 g 235 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 100; DB 26; Length 1015;
Best Local Similarity 65.2%; Pred. No. 5,45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 528 ACCCTGACACGAGATGTATGT 550
||:||||:||||:||||:||||:
QY 1 ACNTYCAICARAAAGGNAAYGT 23

RESULT 8

LOCUS MUSN23K 1047 bp mRNA ROD 14-MAR-1997
DEFINITION Mouse N23K mRNA for N23K protein, complete cds.
ACCESSION D50056
NID g1857023
KEYWORDS N23K protein; N23K.
SOURCE Mus musculus neuroblastoma cell_line:NS20Y cDNA to mRNA.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.
1 (bases 1 to 1047)
Saito,Y., Maruyama,K., Saito,T.C. and Kawashima,S.
N23K, a gene transiently up-regulated during neural
differentiation, encodes a precursor protein for a newly identified
neuropeptide nociceptin

REFERENCE
MEDLINE Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)
96106851
REFERENCE 2 (bases 1 to 1047)
Saito,Y.
TITLE Direct Submission

JOURNAL Submitted (05-APR-1995) to the DBJ/EMBL/GenBank databases. Yumiko
Saito, Tokyo Metropolitan Institute for Medical Science, molecular
biology; Honkomagome 3-18-22, Bunkyo-Ku, Tokyo 113, Japan
(E-mail:saito@rins.hoken.or.jp, Tel:03-3823-2101(ex.5248),
Fax:03-5685-6609)
On Mar 1, 1997 this sequence version replaced gi:1217905.
COMMENT Sequence updated (06-Mar-1996) by:Yumiko Saito.
JOURNAL location/Qualifiers

FEATURES
source

1..1047
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="NS20Y"
/cell_type="neuroblastoma"
57..620
/gene="N23K"
57..620
/gene="N23K"
/note="neuropeptide precursor"
/codon_start=1
/product="N23K protein"
/db_xref="PID:d100935"
/db_xref="PID:g1217906"
/translation="MKILICDILLSLVSSVSSCPDCITCQEKIHPADSEFLKTC
ILCEEKVFPSPILMTCTKVMASWSQSPADPELYSAALYQPRASEMOHLRMRVR
SLVQVDAEPGADAPGADADAEVQKLOKRFGEFTGARKSARKLNOKRFS
SEFMROYLYLSMOSSORRRLRHONGNV"
1040..1045

polyA_signal 1040..1045
BASE COUNT 237 a 315 c 263 g 232 t
ORIGIN

Query Match 100.0%; Score 100; DB 29; Length 1047;
Best Local Similarity 65.2%; Pred. No. 5,45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 594 ACCCTGACACGAGATGTATGT 616
||:||||:||||:||||:||||:
QY 1 ACNTYCAICARAAAGGNAAYGT 23

RESULT 9
LOCUS HSU48263 1198 bp mRNA PRI 23-AUG-1996
DEFINITION Human pre-pro-orphanin FQ (OPF) mRNA, complete cds.
ACCESSION U48263
NID g1185009
KEYWORDS human.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1198)
AUTHORS Nothacker,H.P., Reinscheld,R.K., Mansour,A., Henningsen,R.A.,
Ardatt,A., Monsma,F.J., Jr., Watson,S.V., and Civeilli,O.


```

TITLE Primary structure and tissue distribution of the orphanin FQ precursor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)
MEDLINE 96323283
REFERENCE 2 (bases 1 to 1198)
AUTHORS Nothacker,H.-P. and Henningsen,R.A.
JOURNAL Direct Submission
Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202, Hofmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland
FEATURES
source Location/Qualifiers
1..1198 /organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
1..211
212..742 /gene="OFo"
212..742 /gene="OFo"
CDS /gene="OFo"
/codon_start=1
/product="Pre-pro-orphanin FQ"
/db_xref="PID:91185010"
/translation="MKVLLCDLILSLFSVSFCORCLTCOEKHPALDPSFLVCQ
IIEEEKVEFSPPLMTPTCKVARSSMOLSPAPEYVAALTOPASEGHRLRMDPRVR
SLFOEERPERGMEAGEMEKQLOKREGGTGARAKSLANOKRSFEFRQLVLSS
MOSSORRTTLNGNV"
212..268
/gene="OFo"
/gene="OFo"
599..649 /gene="OFo"
/product="Orphanin FQ"
743..1198
1198 /note="29 A nucleotides"
BASE COUNT 244 a 335 c 345 g 273 t 1 others
ORIGIN
Query Match 100.0%; Score 100; DB 28; Length 1198;
Best Local Similarity 65.2%; Pred. No. 5,45e+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Db 716 ACCCTGCACACAGATGGTAATGT 738
Oy 1 ACNTNCAICARAPAYGNNAATGT 23
RESULT LOCUS MMPNEX3 1253 bp DNA ROD 10-SEP-1996
DEFINITION M.musculus gene encoding prepronociceptin, exon 3.
ACCESSION X93732
ID 91532038
KEYWORDS prepronociceptin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondria eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1253)
Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G., Weunier,J.C. and Parmentier,M.
Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)
96323281
2 (bases 1 to 1253)
Parmentier,M.
Direct Submission
Submitted (19-APR-1996) M. Parmentier, Université Libre de Bruxelles, I R I B N ULB Campus Erasme, 808 Route de Lennik, B-1070 Bruxelles, BELGIUM
JOURNAL Location/Qualifiers
1..1253 /organism="Mus musculus"

```

```

/strain="129/Sv"
/db_xref="taxon:10090"
/clone_1id="lambda fixII (Stratagene #946305)"
/clone="MG43"
605..1087
/number=3
/label=ex3
/usedin=X9/371:MMPN_CDS
/usedin=X9/371:MMPN_mRNA
/product="prepronociceptin"
BASE COUNT      356 a      307 c      323 g      267 t
ORIGIN

Query Match      100.0%; Score 100; DB 29; Length 1253;
Best Local Similarity 65.2%; Pred. No. 5,456+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 1016 ACCCGCACCGAATGCTAATGT 1038
||:||||:||||:||||:||||:
QY 1 ACNTYTCAYCARAAATGNAATGT 23

RESULT 11
LOCUS      MUSNOP      1354 bp      mRNA      ROD      09-APR-1997
DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
ACCESSION  D82866
            g1311472
NID      nociceptin/orphanin FQ.
KEYWORDS  Mus musculus adult brain cDNA to mRNA.
SOURCE    Mus musculus
ORGANISM  Mus musculus
           Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;
           Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
           Murinae; Mus.
           1 (bases 1 to 1354)
REFERENCE  1 Direct Submission
           Submitted (23-DEC-1995) to the DDBJ/EMBL/GenBank databases. Hiroshi
           Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of
           Medicine, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
           (E-mail:takeshimem.u-tokyo.ac.jp, Tel:03-3812-2111(ex.3422),
           Fax:03-3815-9360)
           2 (bases 1 to 1354)
           Takeshima,H.
           Unpublished (1996)
REFERENCE  3 (sites)
           Houtani,T., Nishii,M., Takeshima,H., Nukada,T. and Sugimoto,T.
           Structure and regional distribution of nociceptin/orphanin FQ
           precursor
           Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)
           96216718
JOURNAL    MEDLINE
FEATURES
SOURCE
Location/Qualifiers
1..1354
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="adult"
/tissue_type="brain"
360..923
/codon_start=1
/product="nociceptin/orphanin FQ"
/db_xref="PID:g13114281"
/db_xref="PID:g1311473"
/translaton="MKILFCDVLLSLSSVFSQPRDCLKOEKLPAPDSFNLKTC"
ILOEEVFPPEPLWTCVTKWASGSGOLSPADPELVSAALYQPAASMOHLKRPVRYR
SLVQVRAEGADAPGADAPGADADPADEVDQKLOKRFPGFTGARSARKLANQKRF
SEFNRQILVSMOSSQRRLIHQNGV"
1317..1322
polya_signal
BASE COUNT      309 a      370 c      386 g      289 t
ORIGIN

Query Match      100.0%; Score 100; DB 29; Length 1354;
Best Local Similarity 65.2%; Pred. No. 5,456+00;
Matches 15; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```

JOURNAL
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 19a D-82152 Martinsried, FRG; E-mail: Mewes@mpis.emblnet.org
3 (bases 1 to 2258)

REFERENCE
Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Bactlet,M.C., Barthe,C., Baur,A., Becam,A.M., Billeau,N., Bolcs,E., Brandt,T., Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C., Contreras,R., Crouzet,M., Cziepluch,C., Demolis,N., Delaveau,T., Doignon,F., Domdey,H., Duesterhaus,S., Dubois,E., Dujon,B., El Bakoury,M., Enliat,K.D., Felsenmann,M., Fiers,W., Fobo,G.M., Fritz,C., Gassenhuber,H., Glensdorff,N., Goifeau,A., Grivell,L.A., de Haan,M., Hain,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K., Jacq,C., Jacquet,M., Jauniaux,J.C., Joniaux,J.L., Kalesse,T., Klesau,C., Kirchuth,L., Koetter,P., Korol,S., Liebl,S., Logghe,M., Lohan,A.J.E., Louis,E.U., Li,Z.Y., Mat,M.J., Mallet,L., Manhaupt,G., Messenguy,F., Miosga,T., Moleman,F., Mueller,S., Nasti,F., Obermaler,B., Perea,U., Pierard,A., Piravandi,E., Pohl,F.M., Poll,T.M., Potier,S., Proft,M., Purrelle,B., Ramzani,Rad,M., Rieger,M., Rose,M., Schaaff-Gerstenschaefer,T., Scherens,B., Schwarlose,C., Skala,U., Slonimski,P.P., Smits,P.H.M., Souciet,J.L., Steensma,H.T., Stucka,R., Driesterau,A., van der Aart,Q.J., van Dyck,L., Vassarotti,A., Vetter,I., Verenderdeels,F., Vissers,S., Wagner,G., de Wierfosse,P., Wolffe,K.H., Zagulski,M., Zimmermann,F.K., Mewes,H.W. and Kline,K. Complete DNA sequence of yeast chromosome II
EMBO J. 13 (24), 5795-5809 (1994)
95112788

TITLE
JOURNAL
MEDLINE
FEATURES
Source
Location/Qualifiers
1. .2258
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="II"
307. .2031
/gene="RFT1"
307. .2031
/gene="RFT1"
/note="ORF YBL020w"
/codon_start=1
/db_xref="PID:g536015"
/db_xref="SWISS-PROT:P38206"
/translation="MAKNSQLPSTSQILERSSTGATFLMGQLFKLVTFINNLLIRSLPRIFGTALFLFIQGVLFPSADAIRLSTLRASDSNGIIDDDEEYOETHYKSLVLAIVNFAVPIEPIWIGPLSIGLAMOYRNINAYFTLPPFRMSIFLWLSIIVEKLSLPEFTVNOFLMVAARSFEESIAVTGCIYVFIYVAOQSRPMGYVTSIDKEGIALAFALGILASITILACIYVMDYIKNPKKLESTRLIKTRNNELKKRPPSTSYEPNDIIOHEKKVYFOCFKRLKLEGGKILNCTCYBEQGITLWLSYGSLTRLLRAPTESLRPLARLLSHNRPNKLSEVLNLTFRFYISLMTIVGPNASSFLHQFLIGKSWSTSVLDITIRVYCFPIPLSLNNGIEAFQSVAGDQILKTSFPMNARS GIPFLNSWMLIEKIKLSTIEGILISNIINMVLRIYICVFLPKRHELFSTSPFFNKKDFKVIILAGSTICLLDMWFLIGYVNLQOPVVNLFAMGLALLIVKRGQIQQSPINRR AVNSKRV"

BASE COUNT
655 a 416 c 441 g 746 t

ORIGIN
Query Match 90.0%; Score 90; DB 24; Length 2258;
Best Local Similarity 60.9%; Pred. No. 6.60e+01;
Matches 14; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 396 ACTTTTCACCAACTGTAACGT 418
||:||||:||||:||||:||||:
QY 1 ACNTTCAYCARAAVGGNAAYGT 23

RESULT 14
LOCUS SCU15087 2551 bp DNA PLN 31-JAN-1996
DEFINITION Saccharomyces cerevisiae nuclear division (RFT1) gene, complete cds.
ACCESSION U15087
KEYWORDS 9558439
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae

Eukaryotes: mitochondrial eukaryotes; Fungi: Ascomycota: Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE
1 (bases 1 to 2551)
Koerte, A., Chong, T., Li, X., Mahane, K. and Cal, M.
JOURNAL
Suppression of the yeast mutation rft1-1 by human p53
J Biol Chem. 270 (38), 22556-22564 (1995)
MEDLINE
95403462
REFERENCE
2 (bases 1 to 2551)
Koerte, A.
AUTHORS
Direct Submission
TITLE
Submitted (23-SEP-1994) Andreas Koerte, Institute of Molecular and Cellular Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore, 0511, Singapore
JOURNAL
Location/Qualifiers
1. 2551
/organism="Saccharomyces cerevisiae"
/strain="YMA2"
/isolate="285"
/db_xref="taxon:4932"
/chromosome="11"
/map="2L"
304..2028
/gene="RFT1"
/db_xref="SGD:L0001626"
304..2028
/note="Identical to hypothetical 66.2 kDa protein encoded by HAP3-PEP1 intergenic region, ORF YBL020w product, Swiss-Prot Accession Number P38206"
/codon_start=1
/db_xref="SGD:L0001626"
/function="nuclear division"
/product="Rft1p"
/db_xref="PID:9558440"
/translation="MAKNSQLPSTSEQLERSTGATFLMAGOLFRTLYVFIINLLIRFSPRIGTAFLEFIOGVLPFSRDAMFLSTRISDSGNGIIDDDEEYQETHYKSKYLOAVNFAVLPFWIGFPLSIGLIMAYRNINAVFTLPFRMSIFLMLSTIVELSSEFFIVNQFLMAYARSRESIAVTGCVNIVYAVQOSRYPMGVVTSIDKEGIALAFALGRLASHITLACYWDLKNEPKKFSRLTKIKRENNELKGPKS TSFFONDILQHFKKVYFOLCFKHLLEGDKLIINSCTVEEOGIYALLSVGSLTRLLFAPIESRLFLARLSHNPKNLKSIEVLVNLFRFYIYLSIMITIVFSPANSFLLQFLGSKWSTSVLDTRYVCFTYFPLSLNGIFAFQSVATGQIILKSHFMAFS GFPLNSWLLERKLESTEGILSLINMNVNLICGVLFKNEHRELFSDSFFENK DFTYIINGSTICLDMMWIFGVKNLQGFVNVLFAMGLALLIVKERQTIOSFINRK AVNSKDY"

BASE COUNT 745 a 474 c 492 g 840 t
ORIGIN

Query Match 90.0%; Score 90; DB 24; Length 2551;
Best Local Similarity 60.9%; Pred. No. 6.60e+01;
Matches 14; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 393 ACTTTTCACCAACGCTGTAAGT 415
Oy 1 ACNTYNCACARAAAGGNAAGT 23

RESULT 15
LOCUS YSCAP 3061 bp DNA PLN 23-NOV-1994
DEFINITION S.cerevisiae HAP3 locus encoding 2 genes which regulate respiratory function, complete cds and 5' end respectively.
ACCESSION M20318
M20318
NID 9577522
KEYWORDS respiratory regulation.
SOURCE S.cerevisiae (strain BMG1-7a) DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryotes; mitochondrial eukaryotes; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 3061)
Hahn, S., Pinkham, J., Wei, R., Miller, R. and Guarente, L.
AUTHORS
The HAP3 regulatory locus of Saccharomyces cerevisiae encodes

divergent overlapping transcripts
Mol. Cell. Biol. 8 (2), 653-663 (1988)
JOURNAL
88174707
MEDLINE
On Nov 28, 1994 this sequence version replaced gi:171642.
COMMENT
Drift entry and computer-readable sequence for [1] kindly provided by S. Hahn, 14-MR-1989

FEATURES
SOURCE
1. 3061
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
complement(<1..1285)
/note="ORF"
/codon_start=1
/db_xref="PID:9577523"
/translation="MAKNSQLPSTSEQLERSTGATFLMAGOLFRTLYVFIINLLIRFSPRIGTAFLEFIOGVLPFSRDAMFLSTRISDSGNGIIDDDEEYQETHYKSKYLOAVNFAVLPFWIGFPLSIGLIMAYRNINAVFTLPFRMSIFLMLSTIVELSSEFFIVNQFLMAYARSRESIAVTGCVNIVYAVQOSRYPMGVVTSIDKEGIALAFALGRLASHITLACYWDLKNEPKKFSRLTKIKRENNELKGPKS TSFFONDILQHFKKVYFOLCFKHLLEGDKLIINSCTVEEOGIYALLSVGSLTRLLFAPIESRLFLARLSHNPKNLKSIEVLVNLFRFYIYLSIMITIVFSPANSFLLQFLGSKWSTSVLDTRYVCFTYFPLSLNGIFAFQSVATGQIILKSHFMAFS GFPLNSWLLERKLESTEGILSLINMNVNLICGVLFKNEHRELFSDSFFENK DFTYIINGSTICLDMMWIFGVKNLQGFVNVLFAMGLALLIVKERQTIOSFINRK AVNSKDY"

BASE COUNT 978 a 613 c 676 g 794 t
ORIGIN

Query Match 90.0%; Score 90; DB 24; Length 3061;
Best Local Similarity 60.9%; Pred. No. 6.60e+01;
Matches 14; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 1174 ACCTTACAGCTTGTGAAGT 1196
Cp 23 ACCTTNCCTTGTGNAAGT 1

Search completed: Tue Apr 27 10:32:45 1999
Job time : 105 secs.

THIS PAGE BLANK (USPTO)

 NWSENF
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:25:48 1999; Maspar time 3.31 Seconds
 Tabular output not generated. 90.655 Million cell updates/sec

Title: >US-09-011-797-4
 Description: (1-8) from US09011797.pep
 Perfect Score: 53
 Sequence: 1 TLHONGNV 8

Scoring table: PAM 150
 Gap 15

Searched: 116695 segs, 37453910 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: plr58
 1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 19.697; Variance 20.848; scale 0.945

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	53	100.0	109	2	S70631	1.31e-02
2	53	100.0	168	2	S60337	1.31e-02
3	53	100.0	176	2	JC6152	1.31e-02
4	53	100.0	181	2	JC6151	1.31e-02
5	53	100.0	187	2	JC4502	1.31e-02
6	45	84.9	1051	2	A8373	1.54e+00
7	44	83.0	384	2	B64960	2.71e+00
8	43	81.1	251	2	F64559	4.73e+00
9	43	81.1	636	2	JC4960	4.73e+00
10	41	77.4	112	2	E71313	1.40e+01
11	41	77.4	224	2	S20463	1.40e+01
12	41	77.4	261	2	B64783	1.40e+01
13	41	77.4	508	2	S54021	1.40e+01
14	41	77.4	726	2	S22258	1.40e+01
15	41	77.4	863	2	S41984	1.40e+01
16	41	75.5	240	2	S71458	2.37e+01
17	40	75.5	240	2	S62088	2.37e+01
18	40	75.5	250	2	S70854	2.37e+01
19	40	75.5	298	2	A64922	2.37e+01
20	40	75.5	328	2	D69452	2.37e+01
21	40	75.5	639	2	A39135	2.37e+01
22	40	75.5	745	2	B70017	2.37e+01
23	39	73.6	74	2	S72807	3.98e+01

RESULT	ENTRY	ALIGNMENTS
24	39	73.6 147 2 B48402
25	39	73.6 149 2 E41317
26	39	73.6 158 2 S34612
27	39	73.6 219 2 B47152
28	39	73.6 232 2 S47846
29	39	73.6 398 2 E71539
30	39	73.6 412 2 C64736
31	39	73.6 439 2 F69281
32	39	73.6 462 2 D32057
33	39	73.6 720 2 S61143
34	39	73.6 723 2 A36481
35	39	73.6 885 2 I38968
36	39	73.6 893 2 S63378
37	39	73.6 971 1 JQ1634
38	38	71.7 238 2 D69021
39	38	71.7 325 2 F70440
40	38	71.7 533 2 JC1103
41	38	71.7 601 2 E71435
42	38	71.7 634 2 C64454
43	38	71.7 893 2 S46442
44	38	71.7 941 2 A55195
45	38	71.7 1319 2 C43735

RESULT 1
 ENTRY S70631 #type fragment
 TITLE orphanin FQ/nociceptin - mouse (fragment)
 ORGNISM #formal_name Mus musculus #common_name house mouse
 DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997

ACCESSIONS S70631
 REFERENCE S70631
 #authors Pan, Y.X.; Xu, J.; Pasternak, G.W.
 #journal Biochem. J. (1996) 315:11-13
 #title Cloning and expression of a cDNA encoding a mouse brain orphanin FQ/nociceptin precursor.
 #accession S70631
 #status preliminary
 #molecule_type mRNA
 #residues 1-109 #label PAN
 #cross-references EMBL:U44027; NID:91335869; PID:91335870

SUMMARY
 Query Match 100.0%; Score 53; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 1.31e-02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	102 TLHONGNV 109
Qy	1 TLHONGNV 8

RESULT 2
 ENTRY S60337 #type fragment
 TITLE oploild receptor-like ORL(1) receptor antagonist - rat (fragment)
 ORGNISM #formal_name Rattus norvegicus #common_name Norway rat
 DATE 06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change 18-Mar-1997

ACCESSIONS S60337
 REFERENCE S60337
 #authors Meunier, J.C.; Mollereau, C.; Toll, L.; Snaudieu, C.; Ferrante, P.; Monserrat, B.; Butour, J.L.; Guillemot, J.C.; Parmentier, M.; Costentin, J.
 #journal Nature (1995) 377:532-535
 #title Isolation and structure of the endogenous agonist of oploild receptor-like ORL(1) receptor.

#accession S60337
 #status preliminary; not compared with conceptual translation
 #molecule_type mRNA

##residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 53; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 161 TLHONGNV 168
|||||||
QY 1 TLHONGNV 8

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
ACCESSIONS JC6152
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6152
#molecule_type mRNA
#residues 1-176 ##label NOT
#cross-references GB:U48263; NID:g1185009; PID:g1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opioid receptors. It functions in development.
KEYWORDS opioid peptide
SUMMARY #length 176 #molecular-weight 20295 #checksum 498

Query Match 100.0%; Score 53; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 169 TLHONGNV 176
|||||||
QY 1 TLHONGNV 8

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
ACCESSIONS JC6151
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen, R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli, O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ precursor.
#accession JC6151
#molecule_type mRNA
#residues 1-181 ##label NOT
#cross-references GB:U48262; NID:g1185011; PID:g1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a G-protein-coupled receptor sequentially related to the opioid receptors. It is involved in development.
KEYWORDS opioid peptide
SUMMARY #length 181 #molecular-weight 20478 #checksum 3303

Query Match 100.0%; Score 53; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 174 TLHONGNV 181
|||||||
QY 1 TLHONGNV 8

RESULT 5
ENTRY JC4502 #type complete
TITLE opioid-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N23K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 10-Sep-1997
ACCESSIONS JC4502
REFERENCE JC4502
#authors Salto, Y.; Maruyama, K.; Saido, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label SAT
#cross-references DBJ:D50056; NID:g1857023; PID:d1009395; PID:g1217906
REFERENCE JC4502
#authors Houtani, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto, T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ precursor.
#accession JC4652
#molecule_type mRNA
#residues 1-187 ##label HOI
#cross-references DBJ:D82866; NID:g1311472; PID:d1012281; PID:g1311473
#experimental_source brain
#note The authors translated the codon CTC for residue 72 as lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions as a neuropeptide precursor and plays a role in neuronal differentiation and development.
GENETICS N23K
#gene brain; differentiation; G protein-coupled receptor; neuropeptide
KEYWORDS
FEATURE
1-18
19-187
109-126 #domain signal sequence #status predicted #label SIG
141-157 #product neuropeptide precursor, N23K #status predicted
160-176 #region D-A-E-P-G-A motif
SUMMARY #label MAT
#product neuropeptide #status predicted #label NRP
#product heptadecapeptide #status predicted #label HRP
#length 187 #molecular-weight 20864 #checksum 8738

Query Match 100.0%; Score 53; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.31e-02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 180 TLHONGNV 187
|||||||
QY 1 TLHONGNV 8

RESULT 6
ENTRY A38373 #type complete
TITLE ubiquitin--protein ligase (EC 6.3.2.19) E1 - wheat
ALTERNATE_NAMES ubiquitin-activating enzyme E1, UBA1
ORGANISM #formal_name Triticum sp. #common_name wheat
DATE 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Jul-1998
ACCESSIONS A38373; A42873

```
REFERENCE A38373
#authors Hatfield, P.M.; Callis, J.; Vester, R.D.
#journal J. Biol. Chem. (1990) 265:15813-15817
#title Cloning of ubiquitin activating enzyme from wheat and
#expression of a functional protein in Escherichia coli.
#cross-references NID:90368797
#accession A38373
#status preliminary
#molecule_type mRNA
#residues 1-1051 #label HAT
#cross-references GB:M55604; GB:M33631; NID:g170779; PID:g170780
REFERENCE A42873
#authors Hatfield, P.M.; Vester, R.D.
#journal J. Biol. Chem. (1992) 267:14799-14803
#title Multiple forms of ubiquitin-activating enzyme E1 from wheat.
#identification of an essential cysteine by in vitro
#mutagenesis.
#cross-references NID:92340519
#accession A42873
#status preliminary: not compared with conceptual translation
#molecule_type mRNA
#residues 1-1051 #label HAT
#note sequence extracted from NCBI backbone (NCBIP:109415)
CLASSIFICATION #superfamily ubiquitin-activating enzyme E1
KEYWORDS ligase
SUMMARY #length 1051 #molecular-weight 117007 #checksum 9170

Query Match 84.9%; Score 45; DB 2; Length 1051;
Best Local Similarity 75.0%; Pred. No. 1.54e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 93 TLHDDGCV 100
OY 1 TLHONGNV 8

RESULT 7
ENTRY B64940 #type complete
TITLE hypothetical protein b1794 - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997
ACCESSION B64940
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references NID:97426517
#accession B64940
#status preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-384 #label BLAT
#cross-references GB:AE000274; GB:U00096; NID:g1788089; PID:g1788095;
UMGP:b1794
SUMMARY #experimental_source strain K-12, substrain MG1655
#length 384 #molecular-weight 43628 #checksum 3691

Query Match 83.0%; Score 44; DB 2; Length 384;
Best Local Similarity 75.0%; Pred. No. 2.71e+00;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 284 TLQONGCV 291
OY 1 TLHONGNV 8

RESULT 8
ENTRY F64559 #type complete
```

```
TITLE conserved hypothetical protein HP0318 - Helicobacter pylori
#formal_name Helicobacter pylori
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
05-Jun-1998
ACCESSION F64559
REFERENCE A64520
#authors Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klein,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, J.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Waidman, J.M.;
Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references NID:97394467
#accession F64559
#status preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-251 #label TOM
#cross-references GB:AE000550; GB:AE000511; NID:g2313417; PID:g2313418;
TIGR:HP0318
CLASSIFICATION #superfamily conserved hypothetical protein HP0318
SUMMARY #length 251 #molecular-weight 28507 #checksum 8065

Query Match 81.1%; Score 43; DB 2; Length 251;
Best Local Similarity 75.0%; Pred. No. 4.73e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 105 TLHPNGCV 112
OY 1 TLHONGNV 8

RESULT 9
ENTRY JC4960 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) B -
Thermotoga maritima
ALTERNATE_NAMES DNA gyrase B like; type II DNA topoisomerase
ORGANISM #formal_name Thermotoga maritima
DATE 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change
08-Sep-1997
ACCESSION JC4960
REFERENCE JC4960
#authors Guipaud, O.; Labedan, B.; Forterre, P.
#journal Gene (1996) 174:121-128
#title A gyb-like gene from the hyperthermophilic bacterium
Thermotoga maritima.
#accession JC4960
#molecule_type DNA
#residues 1-636 #label GVI
#cross-references GB:U49692; NID:g1622790; PID:g1622792
COMMENT This enzyme is an adenosine triphosphate dependent enzyme with DNA
relaxation and decatenation activities.
GENETICS
top2B
CLASSIFICATION #superfamily DNA topoisomerase (ATP-hydrolyzing) chain B
KEYWORDS isomerase
SUMMARY #length 636 #molecular-weight 72459 #checksum 3136

Query Match 81.1%; Score 43; DB 2; Length 636;
Best Local Similarity 62.5%; Pred. No. 4.73e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 60 TLHDDGCV 67
```

QY 1 TLHONGNV 8

RESULT 10

ENTRY E71313 #type complete

TITLE probable V-type ATPase, subunit F - syphilis spirochete

ORGANISM #formal_name Treponema pallidum subsp. pallidum #common_name syphilis spirochete

DATE 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 07-Aug-1998

ACCESSIONS E71313

REFERENCE A71250

#authors Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Petersen, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDonald, L.; Artlisch, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

#journal Science (1998) 281:375-388

#title Complete genome sequence of Treponema pallidum, the syphilis spirochete.

#accession E71313

##status preliminary; nucleic acid sequence not shown;

##molecule_type DNA

##residues 1-112 #label COL

##cross-references GB:AE001228; GB:AE000520; NID:g3322816; PID:g3322823

##experimental_source strain Nichols

GENETICS

SUMMARY TP0531

#length 112 #molecular-weight 12494 #checksum 6378

Query Match 77.4%; Score 41; DB 2; Length 112;

Best Local Similarity 62.5%; Pred. No. 1.40e+01;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 89 ALHOGNV 96

QY 1 TLHONGNV 8

RESULT 11

ENTRY S20463 #type complete

TITLE siderophore biosynthesis regulatory protein sfp - Bacillus subtilis (strain OKB105)

ALTERNATE_NAMES sfp

ORGANISM #formal_name Bacillus subtilis

DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Sep-1998

ACCESSIONS S20463

REFERENCE S20463

#authors Nakano, M.M.; Corbelli, N.; Besson, J.; Zuber, P.

#journal Mol. Gen. Genet. (1992) 232:313-321

#title Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis.

#cross-references MUID:92212296

#accession S20463

##molecule_type DNA

##residues 1-224 #label NAK

##cross-references EMBL:X63158; NID:g40138; PID:g40139

##experimental_source strain OKB105

COMMENT This protein (sfp variant) is found in surfactin producing strains. Strains that do not produce surfactin have a sfp(0) variant with a carboxyl-terminal sequence that differs from this sequence due to a single nucleotide deletion and frameshift. The sequence of a sfp(0) variant can be found in FIR:A36931.

GENETICS

#gene sfp; sfp(0)

CLASSIFICATION #superfamily siderophore biosynthesis regulatory protein sfp

KEYWORDS antibiotic biosynthesis

SUMMARY #length 224 #molecular-weight 26135 #checksum 6728

Query Match 77.4%; Score 41; DB 2; Length 224;

Best Local Similarity 71.4%; Pred. No. 1.40e+01;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 172 LHQDQV 178

QY 2 LHONGNV 8

RESULT 12

ENTRY B64783 #type complete

TITLE hypothetical protein b0515 - Escherichia coli

ORGANISM #formal_name Escherichia coli

DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Jul-1998

ACCESSIONS B64783

REFERENCE A64720

#authors Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

#journal Science (1997) 277:1453-1462

#title The complete genome sequence of Escherichia coli K-12.

#cross-references MUID:97426517

#accession B64783

##status nucleic acid sequence not shown; translation not shown

##molecule_type DNA

##residues 1-261 #label BLAT

##cross-references GB:AE000157; GB:U00096; NID:g1786716; PID:g1786725; UMG:P:00515

##experimental_source strain K-12, substrain MG1655

SUMMARY #length 261 #molecular-weight 28730 #checksum 3167

Query Match 77.4%; Score 41; DB 2; Length 261;

Best Local Similarity 100.0%; Pred. No. 1.40e+01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 65 TLHONG 70

QY 1 TLHONG 6

RESULT 13

ENTRY S54021 #type complete

TITLE FMS1 protein - yeast (Saccharomyces cerevisiae)

ALTERNATE_NAMES protein YW9711.09; protein YMR020W

ORGANISM #formal_name Saccharomyces cerevisiae

DATE 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1998

ACCESSIONS S54021; S67323; S72321

REFERENCE S54014

#authors Lye, G.; Churcher, C.M.

#submission submitted to the EMBL Data Library, May 1995

#accession S54021

##molecule_type DNA

##residues 1-508 #label LYE

##cross-references EMBL:Z49211; NID:g798922; PID:g798930; MIPS:YMR020W

##experimental_source strain AB972

REFERENCE S67323

#authors Pousset, D.; Marcilieu, M.; Karst, F.

#submission submitted to the EMBL Data Library, September 1994

#description Characterisation of FMS1, a multicopy suppressor of farnopropion resistance in yeast.

#accession S67323

##molecule_type DNA

##residues 1-508 #label POU

##cross-references EMBL:X81848; NID:g1143555; PID:g113324; PID:g1143556

REFERENCE S72321

#authors Joets, J.; Pousset, D.; Marcilieu, C.; Karst, F.

#journal Curr. Genet. (1996) 30:115-120

#title Characterization of the Saccharomyces cerevisiae FMS1 gene
related to Candida albicans corticosteroid-binding protein

#accession S72321
#status nucleic acid sequence not shown
#molecule-type DNA
#residues 1-508 ##label JOE
##cross-references EMBL:X81848; NID:g1143555; PID:e119324; PID:g1143556

GENETICS
#gene SGD:FMS1
##cross-references SGD:S0004622; MIPS:YMR020w
#map_position 13R

FUNCTION
#description multicopy suppressor of femp1 morph resistance
#keywords transmembrane protein

FEATURE
11-27 #domain transmembrane #status predicted #label TMM
SUMMARY #length 508 #molecular-weight 57805 #checksum 4542

Query Match 77.4%; Score 41; DB 2; Length 508;
Best Local Similarity 100.0%; Pred. No. 1.40e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 TLHONG 31
1 TLHONG 6

RESULT 14
ENTRY S22258 #type complete
TITLE probable protein kinase KIM82 (EC 2.7.1.-) - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 10-Jul-1998

ACCESSIONS
REFERENCE S22258; S19507; S19769
#authors Wilson, C.; Bergantino, E.; Lanfranchi, G.; Valle, G.;
Carignani, G.; Frontali, L.
#journal Yeast (1992) 8:71-77
#title A putative serine/threonine protein kinase gene on chromosome
III of Saccharomyces cerevisiae.

#cross-references MIMD:92254506
#accession S22258
#molecule-type DNA
#residues 1-726 ##label MIL
##cross-references EMBL:X59720; NID:g1907116; PID:e264713; PID:g1907229

REFERENCE S19504
#authors Frontali, L.; Grisanti, P.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19507
#molecule-type DNA
#residues 1-580 ##label PRO
##cross-references EMBL:X59720; MIPS:YCR091w

REFERENCE S19508
#authors Bergantino, E.; Carignani, G.; Lanfranchi, G.; Valle, G.
#submission submitted to the Protein Sequence Database, March 1992
#accession S19769
#molecule-type DNA
#residues 292-726 ##label BER
##cross-references EMBL:X59720; MIPS:YCR091w

GENETICS
#gene SGD:KIM82
##cross-references SGD:S0000687; MIPS:YCR091w
#map_position 3R

CLASSIFICATION #superfamily unassigned Ser/Thr or Tyr-specific protein
kinases; protein kinase homology
KEYWORDS ATP; phosphotransferase; serine/threonine-specific protein
kinase

FEATURE
322-602 #domain protein kinase homology #label KIM
330-338 #region protein kinase ATP-binding motif\

353 #active_site lys #status predicted
SUMMARY #length 726 #molecular-weight 82122 #checksum 5589

Query Match 77.4%; Score 41; DB 2; Length 726;
Best Local Similarity 71.4%; Pred. No. 1.40e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 457 LHONGHY 463
1 LHONGHY 8

RESULT 15
ENTRY S41984 #type complete
TITLE SIP1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change 06-Feb-1998

ACCESSIONS
REFERENCE S41984; S69705
#authors Yang, X.; Hubbard, E.J.A.; Carlson, M.
#journal Science (1992) 257:680-682
#title A protein kinase substrate identified by the two-hybrid
system.

#accession S41984
#status nucleic acid sequence not shown
#molecule-type DNA
#residues 1-863 ##label YAN
##cross-references EMBL:M90531; NID:g172603; PID:g172604

REFERENCE S69553
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, August 1995
#description The sequence of S. cerevisiae lambda 3641 and cosmids 9461,
9831, and 9410.

#accession S69705
#molecule-type DNA
#residues 1-863 ##label DIE
##cross-references EMBL:U33007; NID:g927685; PID:g927721; MIPS:YDR422C

GENETICS
#gene SGD:SIP1
##cross-references SGD:S0002830; MIPS:YDR422C
#map_position 4R

SUMMARY #length 863 #molecular-weight 96258 #checksum 2426

Query Match 77.4%; Score 41; DB 2; Length 863;
Best Local Similarity 62.5%; Pred. No. 1.40e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 754 TLOOGNI 761
1 TLHONGHY 8

Search completed: Fri Apr 16 14:26:05 1999
Job time: 17 secs.

THIS PAGE BLANK (USPTO)

 M O S E R (TM)

Release 3.1a John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:21:56 1999; MasPar time 3.74 Seconds
 170.260 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 131
 Sequence: 1 FSEFMROYLVLSMOSSQ 17

Scoring table: PAM 150
 GAP 15

Searched: 116695 segs, 37453910 residues

Post-Processing: Minimum Match 0%
 Listing first 45 summaries

Database: p158
 1:plrl 2:plrl2 3:plrl3 4:plrl4

Statistics: Mean 28.933; Variance 47.326; scale 0.611

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	100.0	109	2	S70631	9.72e-13
2	131	100.0	168	2	S60337	9.72e-13
3	131	100.0	176	2	JC6152	9.72e-13
4	131	100.0	181	2	JC6151	9.72e-13
5	131	100.0	187	2	JC4502	9.72e-13
6	68	51.9	210	2	G70523	1.19e+00
7	68	51.9	310	2	D70139	1.19e+00
8	50.4	46.8	242	2	S42370	2.50e+00
9	48.9	46.8	207	2	F64479	5.18e+00
10	48.9	46.8	312	2	G71131	5.18e+00
11	47.3	46.8	164	2	D71472	1.06e+01
12	47.3	46.8	320	2	S66176	1.06e+01
13	47.3	46.8	406	2	E70303	1.06e+01
14	46.6	46.8	318	2	JC6059	1.50e+01
15	46.6	46.8	733	2	S67657	1.50e+01
16	46.6	46.8	734	2	I38080	1.50e+01
17	46.6	46.8	792	2	A70476	1.50e+01
18	46.6	46.8	894	2	S45135	1.50e+01
19	46.6	46.8	994	2	J00151	1.50e+01
20	46.6	46.8	1189	2	JC2366	1.50e+01
21	46.6	46.8	144	2	D69936	2.13e+01
22	45.8	44.3	422	2	B24815	2.13e+01
23	45.8	44.3	435	2	H69607	2.13e+01

24	60	45.8	700	1	C1HRH2	calpain (EC 3.4.22.17	2.13e+01
25	60	45.8	700	2	S38361	calpain (EC 3.4.22.17	2.13e+01
26	60	45.8	1187	1	JC4155	protein-tyrosine-phos	2.13e+01
27	60	45.8	4466	1	S17653	dynamin beta heavy cha	2.13e+01
28	60	45.8	4466	1	S17231	dynamin beta heavy cha	2.13e+01
29	59	45.0	89	2	A47074	lipopolysaccharide co	3.00e+01
30	59	45.0	126	2	A69250	hypothetical protein	3.00e+01
31	59	45.0	261	2	S67623	hypothetical protein	3.00e+01
32	59	45.0	331	2	S30606	senescence-related pr	3.00e+01
33	59	45.0	358	2	S42882	cysteine proteinase (3.00e+01
34	59	45.0	367	2	F64232	hypothetical protein	3.00e+01
35	59	45.0	370	2	S73745	hypothetical protein	3.00e+01
36	59	45.0	434	1	NOHUG	phosphopyruvate hydra	3.00e+01
37	59	45.0	464	1	YKPG	citrate (s1)-synthase	3.00e+01
38	59	45.0	716	2	PC4225	cell division cycle p	3.00e+01
39	59	45.0	923	2	E70820	hypothetical glycyne-	3.00e+01
40	59	45.0	946	2	C65093	[glutamate--ammonia-1	3.00e+01
41	59	45.0	968	2	S47105	myosin heavy chain MY	3.00e+01
42	58	44.3	316	2	S00519	ethylene-forming enzy	4.21e+01
43	58	44.3	319	2	S41395	ethylene-forming enzy	4.21e+01
44	58	44.3	530	2	S57907	serine-type carboxype	4.21e+01
45	58	44.3	1247	1	RDECN4	nitrate reductase (EC	4.21e+01

ALIGNMENTS

RESULT 1
 ENTRY S70631 #type fragment
 TITLE orphanin FQ/nociceptin - mouse (fragment)
 ORGANISM #formal_name Mus musculus #common_name house mouse
 DATE 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 10-Sep-1997

ACCESSIONS S70631
 REFERENCE S70631
 #authors Pan, Y.X.; Xu, J.; Pasternak, G.W.

#journal Blochem. J. (1996) 315:11-13
 #title Cloning and expression of a cDNA encoding a mouse brain
 #accession orphanin FQ/nociceptin precursor.
 S70631

#status preliminary
 #molecule_type mRNA
 #residues 1-109 #label PAN

SUMMARY #cross-references EMBL:U04027; NID:q1335869; PID:q1335870
 #length 109 #checksum 9509

Query Match 100.0%; Score 131; DB 2; Length 109;

Best Local Similarity 100.0%; Pred. No. 9.72e-13;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 82 FSEFMROYLVLSMOSSQ 98
 Oy 1 FSEFMROYLVLSMOSSQ 17

RESULT 2
 ENTRY S60337 #type fragment
 TITLE opiooid receptor-like ORL(1) receptor antagonist - rat
 ORGANISM (fragment)
 DATE 06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change 18-Mar-1997

ACCESSIONS S60337
 REFERENCE S60337
 #authors Meunier, J.C.; Mollereau, C.; Toll, L.; Snaudeau, C.;
 Moland, C.; Alvinerie, P.; Butour, J.L.; Guillemot, J.C.;

#journal Nature (1995) 377:532-535
 #title Isolation and structure of the endogenous agonist of opiooid
 #accession S60337
 #status preliminary; not compared with conceptual translation
 #molecule_type mRNA

```

##residues 1-168 ##label MEU
SUMMARY #length 168 #checksum 3964

Query Match 100.0%; Score 131; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 141 FSEFMROYLVLSMOSSQ 157
OY 1 FSEFMROYLVLSMOSSQ 17

RESULT 3
ENTRY JC6152 #type complete
TITLE orphanin FQ precursor - human
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6152
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession precursor.
#molecule_type mRNA
#residues 1-176 ##label NOT
#cross-references GB:U48263; NID:q1185009; PID:q1185010
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opiod
receptors. It functions in development.
#peptide opiod peptide
#length 176 #molecular-weight 20295 #checksum 498

KEYWORDS
SUMMARY
Query Match 100.0%; Score 131; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 149 FSEFMROYLVLSMOSSQ 165
OY 1 FSEFMROYLVLSMOSSQ 17

RESULT 4
ENTRY JC6151 #type complete
TITLE orphanin FQ precursor - rat
ALTERNATE_NAMES nociceptin precursor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 11-Apr-1997 #sequence_revision 09-May-1997 #text_change
10-Sep-1997

ACCESSIONS JC6151
REFERENCE JC6151
#authors Nothacker, H.P.; Reinscheid, R.K.; Mansour, A.; Henningsen,
R.A.; Ardati, A.; Monsma Jr., F.J.; Watson, S.J.; Civelli,
O.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:8677-8682
#title Primary structure and tissue distribution of the orphanin FQ
#accession precursor.
#molecule_type mRNA
#residues 1-181 ##label NOT
#cross-references GB:U48262; NID:q1185011; PID:q1185012
COMMENT This protein is a neuropeptide that is an endogenous ligand to a
G-protein-coupled receptor sequentially related to the opiod
receptors. It is involved in development.
#peptide opiod peptide
#length 181 #molecular-weight 20478 #checksum 3303

KEYWORDS
SUMMARY
Query Match 100.0%; Score 131; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;

```

```

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 154 FSEFMROYLVLSMOSSQ 170
OY 1 FSEFMROYLVLSMOSSQ 17

RESULT 5
ENTRY JC4502 #type complete
TITLE opiod-receptor like G protein coupled receptor
ALTERNATE_NAMES nociceptin/orphanin FQ protein precursor - mouse
ORGANISM N23K protein; nociceptin precursor homolog
#formal_name Mus musculus #common_name house mouse
DATE 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change
10-Sep-1997

ACCESSIONS JC4502, JC4652
REFERENCE JC4502
#authors Saito, Y.; Maruyama, K.; Saito, T.C.; Kawashima, S.
#journal Biochem. Biophys. Res. Commun. (1995) 217:539-545
#title N23K, a gene transiently up-regulated during neural
differentiation, encodes a precursor protein for a newly
identified neuropeptide nociceptin.
#accession JC4502
#molecule_type mRNA
#residues 1-187 ##label SAI
#cross-references DBJ:D82866; NID:q1857023; PID:d1009395; PID:g1217906
#experimental_source NS20Y neuroblastoma cells
REFERENCE JC4652
#authors Houtant, T.; Nishi, M.; Takeshima, H.; Nukada, T.; Sugimoto,
T.
#journal Biochem. Biophys. Res. Commun. (1996) 219:714-719
#title Structure and regional distribution of nociceptin/orphanin FQ
#accession precursor.
#molecule_type mRNA
#residues 1-187 ##label HOU
#cross-references DBJ:D82866; NID:q1311472; PID:d1012281; PID:g1311473
#experimental_source brain
#note The authors translated the codon CRC for residue 72 as
Lys and GAC for residue 76 as Glu
COMMENT This protein belongs to the opiate precursor family, and functions
as a neuropeptide precursor and plays a role in neuronal
differentiation and development.

GENETICS N23K
#gene brain; differentiation; G protein-coupled receptor;
#keywords neuropeptide
FEATURE 1-18
19-187
109-126 #domain signal sequence #status predicted #label SIG\
141-157 #product neuropeptide precursor, N23K #status predicted
160-176 #label MARV
#region D-A-E-P-G-A motif\
#product neuropeptide #status predicted #label NRP\
#product heptadecapeptide #status predicted #label HPP
SUMMARY #length 187 #molecular-weight 20884 #checksum 8758

Query Match 100.0%; Score 131; DB 2; Length 187;
Best Local Similarity 100.0%; Pred. No. 9.72e-13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 160 FSEFMROYLVLSMOSSQ 176
OY 1 FSEFMROYLVLSMOSSQ 17

RESULT 6
ENTRY G70523 #type complete
TITLE hypothetical protein RV0302 - Mycobacterium tuberculosis
#strain H37/RV
ORGANISM (strain H37/RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS G70523

```

REFERENCE
#anchors A70500
Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigleman, K.; Gas, S.; Barry III, C.E.; Tekale, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holt, S.; Hornsby, L.; Jags, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, S.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#accession M5023
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-210 #label COL
#cross-references GB:296800; GB:AL123456; NID:93261800; PID:9321660; PID:92193923
#experimental_source strain H37Rv
GENETICS
#gene Rv0302
SUMMARY
#length 210 #molecular_weight 23396 #checksum 5675
Query Match 51.9%; Score 68; DB 2; Length 210;
Best Local Similarity 43.8%; Pred. No. 1.19e+00;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
DB 106 PDB:1FRL:FLYLSMERSQ 121
QY 2 SEFMROYLVLSMOSSQ 17
RESULT 7
ENTRY D70139 #type complete
TITLE conserved hypothetical integral membrane protein BB0317 - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
ACCESSIONS D70139
REFERENCE A70100
#authors Fraser, C.M.; Castens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kierulff, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vagt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Matthey, L.; McDonald, L.; Arlisch, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references M5023
#accession D70139
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-310 #label KLE
#cross-references GB:AE001138; GB:AE000783; NID:92688210; PID:92688213; TIGR:BB0317
SUMMARY
#experimental_source strain B31
#length 310 #molecular_weight 35680 #checksum 9426
Query Match 51.9%; Score 68; DB 2; Length 310;
Best Local Similarity 50.0%; Pred. No. 1.19e+00;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 228 F5EFPYVIGLMSG 243
QY 1 F5EFPYVIGLMSG 16
RESULT 8
ENTRY S42370 #type complete
TITLE citrate (sl)-synthase (EC 4.1.3.7), precursor - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 18-Sep-1998
ACCESSIONS S42370
REFERENCE S42368
#authors Smith, A.
#submission submitted to the EMBL Data Library, March 1994
#accession S42370
#status preliminary
#molecule_type DNA
#residues 1-468 #label SMT
#cross-references EMBL:230423; NID:9458479; PID:9458482
GENETICS
#introns 69/3; 202/3; 309/3
CLASSIFICATION #superfamily citrate (sl)-synthase
KEYWORDS carbon-carbon lyase; mitochondrion; oxo-acid-lyase
SUMMARY
#length 468 #molecular_weight 51540 #checksum 2659
Query Match 50.4%; Score 66; DB 2; Length 468;
Best Local Similarity 63.6%; Pred. No. 2.50e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
DB 253 F64479:FLYLV 263
QY 1 F5EFPYVIGLMSG 11
RESULT 9
ENTRY F64479 #type complete
TITLE the mononuclease precursor homolog - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
ACCESSIONS F64479
REFERENCE A64300
#authors Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kierulff, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrman, J.D.; Nguyen, D.; Uterback, T.R.; Kelley, J.M.; Peterson, K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
#cross-references M5023
#accession F64479
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-207 #label BUL
#cross-references GB:U67584; GB:L77117; NID:91592077; PID:91592087; TIGR:MJ1439; PID:91511458
GENETICS
#map_position FOR1406535-1407158
SUMMARY
#length 207 #molecular_weight 24547 #checksum 5813
Query Match 48.9%; Score 64; DB 2; Length 207;
Best Local Similarity 60.0%; Pred. No. 5.18e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 6 IMRKLISM 15
:|||||
QY 4 FMRQYVLVLM 13

RESULT 10
ENTRY 671131 #type complete
TITLE Probable ATP-binding protein - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
G71131
ACCESSIONS 671131
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. (1998) 5:55-76
#journal Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
#title
#accession G71131
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-312 #label KAM
#cross-references GB:AP000003; NID:g3236130; PID:d1030856; PID:g3257230
#experimental_source strain OT3
#note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS
#gene PH0820
SUMMARY #length 312 #molecular-weight 35239 #checksum 6859

Query Match 48.9%; Score 64; DB 2; Length 312;
Best Local Similarity 69.2%; Pred. No. 5.18e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 141 FSKGMROLYVLAM 153
:|||||
QY 1 FSEFMROYVLVLM 13

RESULT 11
ENTRY D71472 #type complete
TITLE Probable thioredoxin disulfide isomerase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 13-Sep-1998
D71472
ACCESSIONS D71472
REFERENCE A71460
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Yatsunov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal submitted to GenBank, May 1998
#description Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
#accession D71472
#status preliminary
#molecule_type DNA
#residues 1-164 #label ARN
#cross-references GB:AE001273; GB:AE001273; NID:g3329238; PID:g3329244
#experimental_source serotype D, strain UW-3/Cx

GENETICS
#gene CT780
SUMMARY #length 164 #molecular-weight 18552 #checksum 4132

Query Match 47.3%; Score 62; DB 2; Length 164;
Best Local Similarity 63.6%; Pred. No. 1.06e+01;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 86 FSEFAKQYICM 96
:|||||
QY 1 FSEFMROYVLV 11

RESULT 12
ENTRY S66176 #type complete
TITLE ACC oxidase (clone AC03) oxidase - muskmelon
ORGANISM #formal_name Cucumis melo- #common_name muskmelon
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 13-Sep-1998
S66176
ACCESSIONS S66176
REFERENCE S66174
#authors Lasserre, E.; Bouguin, T.; Hernandez, J.A.; Bull, J.; Pech, J.C.; Balague, C.
#journal Mol. Gen. Genet. (1996) 251:81-90
#title Structure and expression of three genes encoding ACC oxidase homologs from melon (Cucumis melo L.).
#accession S66176
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-320 #label LAS
#cross-references EMBL:X95553; NID:g2104812; PID:e221407; PID:g1183900
#note the nucleotide sequence was submitted to the EMBL Data Library, February 1995

GENETICS
#introns 36/3; 223/3
CLASSIFICATION #superfamily 1-aminocyclopropane-1-carboxylate oxidase
SUMMARY #length 320 #molecular-weight 36397 #checksum 6287

Query Match 47.3%; Score 62; DB 2; Length 320;
Best Local Similarity 29.4%; Pred. No. 1.06e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 282 FDDYMKLYGLKFOAKE 298
:|||||
QY 1 FSEFMROYVLVMSQSQ 17

RESULT 13
ENTRY E70303 #type complete
TITLE hypothetical protein ag_040 - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
E70303
ACCESSIONS E70303
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Anjaj, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MIMD:98196666
#accession E70303
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-406 #label AOF
#cross-references GB:AE000670; NID:g2982779; PID:g2982788; GB:AE000657
#experimental_source strain VFS

GENETICS
#gene ag_040
SUMMARY #length 406 #molecular-weight 47559 #checksum 1115

Query Match 47.3%; Score 62; DB 2; Length 406;
Best Local Similarity 63.6%; Pred. No. 1.06e+01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

DB      191 FEEFREYTVL 201
OY      1 FSEFROYLV 11

RESULT  14
ENTRY   1
TITLE   1-muinnocyclopropane-1-carboxylic acid oxidase (EC 1.-.-.-) -
        muskmelon
ALTERNATE_NAMES ACC oxidase; ethylene-forming enzyme
ORGANISM  #formal_name Cucumis melo #common_name muskmelon
DATE      31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change
        13-Sep-1998
ACCESSIONS
REFERENCE #authors
        Lasserre, E.; Bouquin, T.; Hernandez, J.A.; Bull, J.; Pech,
        J.C.; Balaque, C.
        Mol. Gen. Genet. (1996) 251:81-90
        Structure and expression of three genes encoding ACC oxidase
        homologs from melon (Cucumis melo L.).
#accession JCE6059
#molecule_type mRNA
#residues 1-318 #label LAS
#cross-references EMBL:X95551; NID:q2104811; PID:e221404; PID:g1183896
REFERENCE #experimental_source seed, clone AC01
        S29395
#authors Balaque, C.; Watson, C.F.; Turner, A.J.; Rouge, P.; Picton,
        S.; Pech, J.C.; Grierson, D.
        Eur. J. Biochem. (1993) 212:27-34
        Isolation of a ripening and wound-induced cDNA from Cucumis
        melo L. encoding a protein with homology to the
        ethylene-forming enzyme.
#accession S29395
#status preliminary
#molecule_type mRNA
#residues 1-318 #label BAL
#cross-references EMBL:X69935; NID:q22662; PID:g22663
COMMENT This enzyme is a rate-limiting enzyme in ethylene biosynthesis. It
        catalyses the last step of ethylene biosynthesis, converting
        1-aminocyclopropane-1-carboxylic acid to ethylene, and regulates
        the rate of ethylene production.
GENETICS
#gene cm-ac01
#introns 36/2; 111/2; 222/3
CLASSIFICATION #superfamily 1-aminocyclopropane-1-carboxylate oxidase
KEYWORDS oxidoreductase; seed; transmembrane protein
FEATURE
        278-298
        39,177,234
SUMMARY #domain transmembrane #status predicted #label TM\
        #binding_site iron (His) #status predicted
        #length 318 #molecular_weight 36127 #checksum 7777
Query Match 46.6%; Score 61; DB 2; Length 318;
Best Local Similarity 29.4%; Pred. No. 1.50e+01;
Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

DB      284 FEDYKLYGVKFOAKE 300
OY      1 FSEFROYLVLMSSQ 17

RESULT  15
ENTRY   1
TITLE   S56767, #type complete
        CDC46 homolog - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE      10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
        04-Sep-1998
ACCESSIONS
REFERENCE S56767
        Kimura, H.; Takizawa, N.; Nozaki, N.; Sugimoto, K.
        Nucleic Acids Res. (1995) 23:2097-2104
        Molecular cloning of cDNA encoding mouse Cdc41 and CDC46
        homologs and characterization of the products: physical
        interaction between Pl(MCM3) and CDC46 proteins.

```

```

#accession S56767
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type mRNA
#residues 1-733 #label KIM
#cross-references EMBL:D26090; NID:g940403; PID:d1005624; PID:g940404
#note the nucleotide sequence was submitted to the EMBL Data
        Library, December 1993
CLASSIFICATION #superfamily cell division control protein CDC47
FEATURE
        327-539
        #domain MCM2 core domain similarity #label MCM
        #length 733 #molecular_weight 82342 #checksum 6080
SUMMARY






Query Match 46.6%; Score 61; DB 2; Length 733;
Best Local Similarity 70.0%; Pred. No. 1.50e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB      35 FKEFLROYRV 44
OY      1 FSEFROYLV 10

```

Search completed: Fri Apr 16 14:22:14 1999
 Job time : 18 secs.

THIS PAGE BLANK (USPTO)

/note="encodes an active 17 amino acid peptide:
FGFTGAKRSARLANQ"

BASE COUNT 85 a 105 c 122 g 60 t

Query Match 98.0%; Score 196; DB 29; Length 372;

Best Local Similarity 56.9%; Pred. No. 5,94e-08;

Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 188 TTGGGGGCTTCACCGGGCCCGGAATCAGCCCGAAGTTGGCCACACG 238

OY 1 TTGGNGGNTTACNGGNGCMGNARMSNGCMGNARNTNGCNAAYCAR 51

RESULT 2
LOCUS MPMPMRNA 633 bp RNA ROD 10-SEP-1996
DEFINITION M.musculus mRNA for prepronociceptin.
ACCESSION X97373
NID 91532039
KEYWORDS prepronociceptin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 633)

AUTHORS Mollerereu,C., Simons,M.J., Soularne,P., Liners,F., Vassart,G.,

Meunier,J.C. and Parmentier,M.

Structure, tissue distribution, and chromosomal localization of the

prepronociceptin gene

Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)

96323281

2 (bases 1 to 633)

Parmentier,M.

Direct Submission

Submitted (19-APR-1996) M. Parmentier, Universite Libre de

Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-

1070 Bruxelles, BELGIUM

Location/Qualifiers

1. 633

/organism="Mus musculus"

/strain="129/Sv"

/db_xref="taxon:10090"

/clone_lib="lambda FIXII (Stratagene #946305)"

/clone="MG43"

<1. 150

/number=2

join(<1. 150, 151. .>633)

join(25. 150, 151. .588)

/codon_start=1

/product="prepronociceptin"

/db_xref="PID:e244874"

/db_xref="PID:91532040"

/translation="MKILFCDVLLSLSVSSFCPRDLCTQERLHPAPDFNLKTC

ILQCEKVFPRPLWTVCTKRVASGSGQSLDPELVSAITQPSAEQHLKRPFR

SLVQVRDAEPGADAPGADAPGADAEVQKQKFGFTGAKRSARLANQKRF

SEFMROYLVLSMOSQRRRLIHONGNV"

25. 69

sig_peptide

mat_peptide

exon

/product="prepronociceptin"

/number=3

BASE COUNT 138 a 194 c 182 g 119 t

ORIGIN

Query Match 98.0%; Score 196; DB 29; Length 633;

Best Local Similarity 56.9%; Pred. No. 5,94e-08;

Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 445 TTGGGGGCTTCACCGGGCCCGGAATCAGCCCGAAGTTGGCCACACG 495

OY 1 TTGGNGGNTTACNGGNGCMGNARMSNGCMGNARNTNGCNAAYCAR 51

RESULT 3
LOCUS HSPPNEX3 878 bp DNA PRI 10-SEP-1996
DEFINITION H.sapiens gene encoding prepronociceptin, exon 3.
ACCESSION X97368
NID 91531986
KEYWORDS prepronociceptin.
SOURCE human.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 878)

Mollerereu,C., Simons,M.J., Soularne,P., Liners,F., Vassart,G.,

Meunier,J.C. and Parmentier,M.

Structure, tissue distribution, and chromosomal localization of the

prepronociceptin gene

Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)

96323281

2 (bases 1 to 878)

Parmentier,M.

Direct Submission

Submitted (19-APR-1996) M. Parmentier, Universite Libre de

Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-

1070 Bruxelles, BELGIUM

Location/Qualifiers

1. 878

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="lambda DASHII (Stratagene #94520)"

/clone="HG91"

/chromosome="8"

/map="P21"

/map="1.9 CR from WI-1172 marker"

73. 524

/number=3

/label=ex3

/usedin=X97367:HSPPN_CDS

/usedin=X97367:HSPPN_mRNA

/product="prepronociceptin"

BASE COUNT 194 a 274 c 231 g 179 t

ORIGIN

Query Match 98.0%; Score 196; DB 26; Length 878;

Best Local Similarity 56.9%; Pred. No. 5,94e-08;

Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 334 TTGGGGGCTTCACCGGGCCCGGAGTCGCGCGAAGTTGGCCATCAG 384

OY 1 TTGGNGGNTTACNGGNGCMGNARMSNGCMGNARNTNGCNAAYCAR 51

RESULT 4

LOCUS 579730 925 bp mRNA ROD 30-JAN-1996

DEFINITION ORL1 receptor agonist precursor-endogenous agonist of opioid

receptor-like ORL1 receptor [rats, brain, mRNA Partial, 925 nt].

579730

ACCESSION 579730

NID 91172238

KEYWORDS Rattus sp. brain.

SOURCE Rattus sp.

ORGANISM Rattus sp.

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 925)

Meunier,J.C., Mollerereu,C., Toll,L., Suardau,C., Moland,C.,

Alvenerie,P., Butour,J.L., Guillemot,J.C., Ferrara,P., Monsarrat,B.

et al

Isolation and structure of the endogenous agonist of opioid

receptor-like ORL1 receptor [see comments]

Nature 377 (6549), 532-535 (1995)

JOURNAL MEDLINE 96013745

REMARK GenBank staff at the National Library of Medicine created this

entry [NCBI glibseq 171655] from the original journal article.

TITLE	N23K, a gene transiently up-regulated during neural differentiation, encodes a precursor protein for a newly identified neuropeptide nociceptin		
JOURNAL	Biochem. Biophys. Res. Commun. 217 (2), 539-545 (1995)		
MEDLINE	96106851		
REFERENCE	2 (bases 1 to 1047)		
AUTHORS	Saito, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-Mar-1995) to the DDBJ/EMBL/GenBank databases. Yumiko Saito, Tokyo Metropolitan Institute for Medical Science, molecular biology; Honkomagome 3-18-22, Bunkyo-ku, Tokyo 113, Japan (E-mail: saito@rinsoken.or.jp, Tel:03-3823-2101(ex.5248), Fax:03-5685-6609)		
COMMENT	On Mar 1, 1997 this sequence version replaced gi:1217905. Sequence updated (06-Mar-1996) by:Yumiko Saito.		
FEATURES	Location/Qualifiers		
source	1..1047		
gene	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/cell_line="NS20Y"		
	/cell_type="neuroblastoma"		
	57..620		
	/gene="N23K"		
	57..620		
	/note="N23K"		
	/note="neuropeptide precursor"		
	/codon_start=1		
	/product="N23K protein"		
	/db_xref="PID:d1009395"		
	/db_xref="PID:g1217906"		
	/translation="MKILFCVLLSLTSYFSSCPDCLTCQEKLRHPSPFLKTC		
	ILQEEKVFPRPLTWCTVKVASSGGLSPDPELVSAALQPRASEQKHKRPVRVA		
	SLVWRDAEPADAPGADAPGADDAEEVQKQLQKRFGGFTGARRSARLANQKRR		
	SEFRRQYLVLSMOSQRRRLHQNGNV"		
BASE COUNT	237 a	315 c	263 g
ORIGIN	232 t		
polya_signal	1040..1045		
Query Match	98.0%: Score 196: DB 29: Length 1047:		
Best Local Similarity	56.9%: Pred. No. 5,94e-08:		
Matches	29: Conservative 11: Mismatches 11: Indels 0: Gaps 0:		
Db	477 TTGGGGCTTCACCGGGCCCGGAATCAGCCCGGAAGTTGGCCACACG 527		
QY	1 TTGGGNGNTTYACGNGCNGMNAARMSNGCMGAARYTGCNAAYCAR 51		
RESULT	10		
LOCUS	HSU48263	1198 bp	mRNA PRI 23-AUG-1996
DEFINITION	Human pre-pro-orphanin FQ (OFO) mRNA, complete cds.		
ACCESSION	U48263		
NID	91185009		
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 1198)		
AUTHORS	Nothacker, H. P., Reinscheid, R. K., Mansour, A., Henningsen, R. A., Ardati, A., Monsma, F. J., Jr., Watson, S. J. and Civelli, O.		
TITLE	Primary structure and tissue distribution of the orphanin FQ precursor		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8677-8682 (1996)		
MEDLINE	96323383		
REFERENCE	2 (bases 1 to 1198)		
AUTHORS	Nothacker, H. P. and Henningsen, R. A.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-FEB-1996) Hans-Peter Nothacker, PRPN 69/202, Hoffmann-La Roche AG, Grenzacherstr 124, Basel, 4070, Switzerland		
FEATURES	Location/Qualifiers		
source	1..1198		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		

FEATURES	Source	Location/Qualifiers
5'UTR	gene	/tissue_type="brain"
gene		1..211
		212..742
		/gene="OFQ"
		212..742
		/gene="OFQ"
		/codon_start=1
		/product="pre-pro-orphantin FQ"
		/db_xref="PID:g1185010"
		/translation="MKVLCDLLILSLFSSVSSCCQDCVCRKLTPTLDSEFLKWC ILDEEEVFPSPMTPTCTKYMARSSWQLSPAEEHYAALYOPASMOHLRMPFVVR SLROEKEPEPGEEMAEQEMQKQKRFSGFTGARKSARLANQKRSEEMROVLVLS MOSSQRRRTIHOXGNV"
		212..268
		/gene="OFQ"
		599..649
		/gene="OFQ"
		/product="orphantin FQ"
		743..1198
		1198
		/note="29 A nucleotides"
BASE COUNT	244 a	335 c 345 g 273 t 1 others
ORIGIN		
Query Match		98.0% Score 196; DB 28; Length 1198;
Best Local Similarity		56.9% Pred. No. 5,946-06;
Matches	29; Conservative	11; Mismatches 11; Indels 0; Gaps 0;
Db	599	TTTGGGGGCTTACCGGGGCGGGAAGTCCGCGGAGTGGCCATCAG 649
Oy	1	TTTGGNGGNTTACGNGCNGNNAHMSNCGNMGNAKRTNGCNATCAR 51
RESULT	11	
LOCUS	AB005251	1209 bp mRNA MAM 09-APR-1998
DEFINITION		Bos taurus mRNA for nociceptin/orphantin FQ precursor, complete cds.
ACCESSION	AB005251	
NID	g3041768	
KEYWORDS		nociceptin/orphantin FQ precursor.
SOURCE		Bos taurus cDNA to mRNA.
ORGANISM		Bos taurus
REFERENCE		Eukaryote: Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
AUTHORS		1 (sites) Okuda-Ashtika, E., Minami, T., Tachibana, S., Yoshihara, Y., Nishitani, Y., Kimura, T. and Ito, S.
TITLE		Nocistatin, a peptide that blocks nociceptin action in pain transmission
JOURNAL		Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Okuda-Ashtika, Kansai Medical University, Department of Medical Chemistry; 10-15 Fumizono, Moriyasu, Osaka, JAPAN, Moriyasu, Osaka 570, Japan (E-mail: ashtika@etaki.kmu.ac.jp, Tel: 06-992-1001, Fax: 06-992-1781)
MEDLINE		2 (bases 1 to 1209)
REFERENCE		98180619
AUTHORS		Okuda-Ashtika, E.
TITLE		Direct Submission
JOURNAL		Submitted (26-JUN-1997) to the DDBJ/EMBL/GenBank databases. Emiko Okuda-Ashtika, Kansai Medical University, Department of Medical Chemistry; 10-15 Fumizono, Moriyasu, Osaka, JAPAN, Moriyasu, Osaka 570, Japan (E-mail: ashtika@etaki.kmu.ac.jp, Tel: 06-992-1001, Fax: 06-992-1781)
FEATURES		
Source		Location/Qualifiers
		1..1209
		/organism="Bos taurus"
		/db_xref="taxon:9913"
		210..740
		/codon_start=1
		/product="nociceptin/orphantin FQ precursor"
		/db_xref="PID:d1026347"
		/db_xref="PID:g3041769"
		/translation="MKILFCDLLILSLFSSVSSCCQDCVCRKLTPTLDSEFLKWC ILDEEEVFPSPMTPTCTKYMARSSWQLSPAEEHYAALYOPASMOHLRMPFVVR SLROEKEPEPGEEMAEQEMQKQKRFSGFTGARKSARLANQKRSEEMROVLVLS MOSSQRRRTIHOXGNV"
		240 a 359 c 341 g 269 t
BASE COUNT	240 a	359 c 341 g 269 t
ORIGIN		

Query Match 98.0%; Score 196; DB 20; Length 1209;
Best Local Similarity 56.9%; Pred. No. 5,94e-08;
Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 597 TTTCGGGGCTTCACCGGGCCCGGAATGCGGAGTTGGCCACACAG 647
1 TTTGGNGNTTTCACNGNCNMGNARMSNGCMGNARTNGCNAAYCAR 51

RESULT 12
LOCUS MAPPNEX3 1253 bp DNA ROD 10-SEP-1996
DEFINITION M.musculus gene encoding prepronociceptin, exon 3.
ACCESSION X97372
NID 91532038
KEYWORDS prepronociceptin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 1253)
Mollereau,C., Simons,M.J., Soularue,P., Liners,F., Vassart,G., Meunier,J.C. and Parmentier,M.
Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene
Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8666-8670 (1996)

JOURNAL MEDLINE 96333281
AUTHORS Parmentier,M.
REFERENCE 2 (bases 1 to 1253)
JOURNAL Direct Submission
TITLE Submitted (19-Apr-1996) M. Parmentier, Universite Libre de Bruxelles, I R I B H N ULB Campus Erasme, 808 Route de Lennik, B-1070 Bruxelles, Belgium
FEATURES
source Location/Qualifiers
1..1253
/organism="Mus musculus"
/strain="129/Sv"
/db_xref="taxon:10090"
/clone_lib="Lambda FIXII (Stratagene #946305)"
/clone="MG43"
605..1087
/number=3
/label=ex3
/usedin="X97371:MAPPN_CDS
/usedin="X97371:MAPPN_mRNA
/product="prepronociceptin"

BASE COUNT 356 a 307 c 323 g 267 t
ORIGIN

Query Match 98.0%; Score 196; DB 29; Length 1253;
Best Local Similarity 56.9%; Pred. No. 5,94e-08;
Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 899 TTTCGGGGCTTCACCGGGCCCGGAATCAGCCCGAGTTGCCACACAG 949
1 TTTGGNGNTTTCACNGNCNMGNARMSNGCMGNARTNGCNAAYCAR 51

RESULT 13
LOCUS MDSNOP 1354 bp mRNA ROD 09-APR-1997
DEFINITION Mouse mRNA for nociceptin/orphanin FQ, complete cds.
ACCESSION D82866
NID 91311472
KEYWORDS nociceptin/orphanin FQ.
SOURCE Mus musculus adult brain cDNA to mRNA.
ORGANISM Mus musculus
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE 1 (bases 1 to 1354)
JOURNAL Takeshima,H.
AUTHORS Direct Submission
TITLE

JOURNAL
Submitted (25-DEC-1995) to the DDBJ/EMBL/Genbank databases. Hiroshi Takeshima, University of Tokyo, Dept. of Pharmacology, Faculty of Medicine; Honjo 7-3-1, Bunkyo-ku, Tokyo 113, Japan (E-mail:takeshim@u-tokyo.ac.jp, Tel:03-3812-2111(ex.3422), Fax:03-3815-9360)

REFERENCE 2 (bases 1 to 1354)
Takeshima,H.
JOURNAL Unpublished (1996)
AUTHORS 3 (sites)
Houtani,T., Nishi,M., Takeshima,H., Nukada,T. and Sugimoto,T.
Structure and regional distribution of nociceptin/orphanin FQ precursor.
JOURNAL Biochem. Biophys. Res. Commun. 219 (3), 714-719 (1996)
MEDLINE 96216718
FEATURES
source Location/Qualifiers
1..1354
/organism="Mus musculus"
/db_xref="taxon:10090"
/dev_stage="adult"
/tissue_type="brain"
360..923
/codon_start=1
/product="nociceptin/orphanin FQ"
/db_xref="PID:d1012281"
/db_xref="PID:g1311473"
/translation="MKILFCDYLLSLSSVSSCPDCLTCOEKLPAPDSFNLKTC ILQCEEVFPPLMTVCTKYMAGSSQLSPADEILYSALQYKASEMOLKMPYR SLVOVRDAEPGADPEGADPEADAEVEQQLKRRGGFGARKSARKLANQRF SEFMROYIVLSMOSQRRRLHNGNV"

BASE COUNT 309 a 370 c 386 g 289 t
ORIGIN

polya_signal 1317 1322

Query Match 98.0%; Score 196; DB 29; Length 1354;
Best Local Similarity 56.9%; Pred. No. 5,94e-08;
Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db 780 TTTCGGGGCTTCACCGGGCCCGGAATCAGCCCGAGTTGCCACACAG 830
1 TTTGGNGNTTTCACNGNCNMGNARMSNGCMGNARTNGCNAAYCAR 51

RESULT 14
LOCUS HPAE000638 9956 bp DNA BCT 06-AUG-1997
DEFINITION Helicobacter pylori section 116 of 134 of the complete genome.
ACCESSION AE000638 AE000511
NID 92314547
KEYWORDS
SOURCE Helicobacter pylori.
ORGANISM Helicobacter pylori
Eubacteria; Proteobacteria; epsilon subdivision; Helicobacter.
REFERENCE 1 (bases 1 to 9956)
Tomb,J.-F., White,O., Kervatage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klein,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khaliq,H.G., Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Uterback,T.R., Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
Bowman,C., Matthey,L., Wallin,E., Hayes,W.S., Borodovsky,M., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.
The complete genome sequence of the gastric pathogen Helicobacter pylori
Nature 388 (6642), 539-547 (1997)

JOURNAL Nature 388 (6642), 539-547 (1997)
MEDLINE 97394467
REFERENCE 2 (bases 1 to 9956)
Tomb,J.-F., White,O., Kervatage,A.R., Clayton,R.A., Sutton,G.G., Fleischmann,R.D., Ketchum,K.A., Klein,H.P., Gill,S., Dougherty,B.A., Nelson,K., Quackenbush,J., Zhou,L., Kirkness,E.F., Peterson,S., Loftus,B., Richardson,D., Dodson,R., Khaliq,H.G., Glodek,A., McKenney,K., Fitzgerald,L.M., Lee,N., Adams,M.D., Hickey,E.K., Berg,D.E., Gocayne,J.D., Uterback,T.R., Peterson,J.D., Kelley,J.M., Cotton,M.D., Weidman,J.M., Fujii,C., Karp,P.D., Smith,H.O., Fraser,C.M. and Venter,J.C.

TITLE
 JOURNAL
 COMMENT
 Bowman, C., Matthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.
 Direct Submission
 Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 Address all correspondence to:

hpdbetlgr.org

or

J. Craig Venter
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD, 20850

Douglas E. Berg is with the Washington University St. Louis School of Medicine, St. Louis, MO 63110.

Eric Wallin is with Stockholm University Department of Biochemistry, Stockholm, Sweden

Hamilton O. Smith is with the Johns Hopkins University School of Medicine, Baltimore, MD, 21205.

Mark Borodovsky and William S. Hayes are with the School of Biology, Georgia Tech, Atlanta, GA, 30332.

Peter D. Karp is with SRI International, Artificial Intelligence Center, Menlo Park, CA, 94025.

All other authors are with The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD, 20850.

Coding potential analysis was performed using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332.

e-mail: mark@amber.gatech.edu

Submission and annotation: Owen White
 e-mail: owhite@tlgr.org

Biological role information, putative identifications, sequence alignments, on-line name and sequence search capability are available at TIGR's World Wide Web site.
 (URL: <http://www.tigr.org/tdb/mbd/hpdbh/hpdbh.html>).
 Location/Qualifiers

FEATURES

source

gene

CDS

CDS

gene

```

1..9956
/organism="Helicobacter pylori"
/strain="26695"
/db_xref="taxon:210"
complement(116..556)
/gene="HP1377"
complement(116..556)
/gene="HP1377"
/feature="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1377"
/db_xref="PID:g2314555"
/translation="MSLRRFCAFLKMGSPMIFDVKADILGFETIHKKRLQKIDEIF
LRNTEENSIVSFTLVNPFARLKYEFEPYPLKLELEGAKSVANINVOPTIE
LSTVNYLADLILNLDKQLMGQVLDNSNKYPHYLRNITLSHTHE"
914..1576
/gene="HP1378"
914..1576
/gene="HP1378"
/feature="HP1378"
/feature="similar to EGAD:40678 percent identity: 25.53;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

```

/product="competence lipoprotein (comL)"
/db_xref="PID:g2314548"
/translation="WRLKFKTFLEITAAIVIGTGCAKKKKKDEYKPAIFWYOGI
LREILFANLEADNYSLSQSEHNSIPVPPAMALAGAHKKKREYVLAIFYOGI
RGTGDNDYLFTELQSHYVAFKNHSQDOFISISYLSLEFTEKYSNRPVVEY
MOIKRILGQNELRAIANVYKKRHKPEGVKLELIDETLEKETPKRSHMPVYLIF
DM"
1618..4125
/gene="HP1379"
1618..4125
/gene="HP1379"
/feature="similar to EGAD:37242 percent identity: 43.94;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ATP-dependent protease (lon)"
/db_xref="PID:g2314549"
/translation="WTEDEPKTLLVEEDTFLVPEMTAPIFLONNASSIKAVAKN
KSLVFIACQKDKNDENAPYDVGYIGSYMEANPNRKYVLEFNGIKGRILPEAKE
NEQGLEQISFTEITDYDKEHQIAVEVLEKVTITLVNSLPPDLIKALEDDP
NRADLIAALHLKRDQVSLFANNRTEQRLDLIDIVIEETKQLOKEIKSKYHOK
MEOTNKEVFLKEOLKOIOKEIGTDKQDELDNOYOKLESIKPFLKEEAFKIKOID
RLSRHADSDSATQNIEMLDVPGQYKALDIKRVBQLDKDRHSIKRPERI
VEYFATMQLLEMRKKRPEKDKTKTGTLICFYGPBGVGTSLANSIARSPVIRIA
LGGLDEVNLEGRHRTYIGSPGRVYQGLIAKKNNPYWIDEIDKVRSGDPASA
LLEIDPEQNTAFRDHYANFSDLSQVLFATANNIDIPAPLRDMERISVSTPN
KEEIAKNYLLIPQLEKHALKPSVEISHEOLKLIETRYRAGVDRLOQTATMKR
VALKLEIDNPQKGRTRKKNKSESDQSEKQSENKDFCVSTPPNLTKEYLE
RMVEIDPIDENKIGIYNGLAMTPVGGDVLEKLEVLKRGSELKTSQVYMESA
IIAFSVVILNDNETLKVPIKPISETDAGKKKKLYNNVADLHLHPGATPKDGP
SAGIMASVMSIICDRATRSYVATGTLTSGEVLPIPGLEKELIAAFKAKIATALI
PVKTERDLEIPAEVRENLIIVAKNIAEVLKTL"
4164..4961
/gene="HP1380"
4164..4961
/gene="HP1380"
/feature="similar to EGAD:8704 percent identity: 30.22;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="prephenate dehydrogenase (tyrA)"
/db_xref="PID:g2314550"
/translation="WGSGLGALQWGRFKSYIGDNHNLAKKALTLGLVDEVCGE
KLECDVFLAIPVEGILGCLKMTSIRKSATITDLSGAKQIINIKRSIKNTAA
HPMCTEFYKASVKGQYENALVTLCDLESGTQVQVIAEIVAGVARKIKMSNE
HDTVAVLSHLPVHYVALANSVLRQNDPEMLISLAGGFRDMSRLSKSSPLMWDIF
KONRNVLEATIKCEKEIVQAKWIEENNDESLAEMMAQANKLQEFM"
5013..5246
/gene="HP1381"
5013..5246
/gene="HP1381"
/feature="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1381"
/db_xref="PID:g2314556"
/translation="MRGAKWGIWIFENDYRLFHMDSMDLPSSQOTTIRDYLRGSIIV
FAIYIITSSHSNAYKTLIASNKRPVLSHLEI"
complement(5305..5697)
/gene="HP1382"
complement(5305..5697)
/feature="HP1382"
/feature="hypothetical protein; identified by Genemark; putative"
/codon_start=1
/transl_table=11
/product="H. pylori predicted coding region HP1382"
/db_xref="PID:g2314557"
/translation="MTNITPEAKNINRHSVLVEEGNRLARKYHQVVEELTIKOG
KTFSSKNAISGFMVYHNDTSLDSYENAKSECFYIINDQNTYPLQEMRCKCKEYER
VEKQVVFKNKNELNELPKYINNAKKT"
5666..6148

```

```

/ gene="HP1383"
5666.6148
CDS

```

```

/gene="HP1383"
/notes="similar to EGAD:44312 percent identity: 38.05
identified by sequence similarity; putative"
(codon start=1

```

```

//transl_table=11
//Product="restriction modification system S subunit"
//db_xref="PID:g2314551"
//translation="MFASGVILVLIKEIATNGFOSLPIEKINDEFYLIITLTK
NKLTLASGTFLEWSPMKIRKRLIPIEPLNMOIAINILSDIRYVINDAAILKEE
SVKRALISELLSQRRKLRKGFQMGAVRLGDIANYIIMPNSKIAEFLGGCKRRLNSPAR
"
complement(6173)..6379)
/gene="HP1384"
complement(6173..6379)

```

```

/gene="HP1384"
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=11
/transl_table=11
/product="H. pylori predicted coding region HP1384"
/db_xref="GeneID:39314559"
translation="MKNSSVLTLEEFERNALLIQAOEKEQVTSICLKKEQVLTNKK"

```

...
Note: remainder of annotations omitted.

Query Match	55.5%;	Score 11;	DB 16;	Length 9956;
Best Local	Similarity 46.5%;	Pred. NO. 2.18e+01;		
Matches	20;	Conservative	8;	Mismatches 15;
			Indels	0;
			Gaps	0;

Db 7520 GCGAGCTTTTTCGAGATTGTTGCTCCTTTAAACCGCAA 7562

Cp 44 GCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCCNCRA 2

RESULT	15
10000000	10000000

ACCESSION	LOCUS	DEFINITION	INVERTED REPEAT	DATE
AC004328	55452 bp	Drosophila melanogaster, chromosome 2R, region 54E8-54F2, pl clone DS02155, complete sequence.	INV	01-DEC-1998
AC004328	AC003872	AC003607	AC003674	

ACCU038/4	ACCU038/3	ACCU038/2	ACCU038/1	ACCU038/0
fruit fly.	fruit fly.	fruit fly.	fruit fly.	fruit fly.
<i>Drosophila melanogaster</i>	<i>Drosophila melanogaster</i>	<i>Drosophila melanogaster</i>	<i>Drosophila melanogaster</i>	<i>Drosophila melanogaster</i>
HTG.	HTG.	HTG.	HTG.	HTG.
g3928064	g3928064	g3928064	g3928064	g3928064
NID	NID	NID	NID	NID
KEYWORDS	KEYWORDS	KEYWORDS	KEYWORDS	KEYWORDS
SOURCE	SOURCE	SOURCE	SOURCE	SOURCE
ORGANISM	ORGANISM	ORGANISM	ORGANISM	ORGANISM
COLLECTION	COLLECTION	COLLECTION	COLLECTION	COLLECTION

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 55452)
AUTHORS
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazet, R.G.,

Butenhoff, C., Champagne, M., Chavez, C., Chew, M., Ciesliko, L., Doyle, C. M., Fafian, D. E., Galle, R., George, R. A., Harris, N. L., Hoskins, R. A., Houston, K. A., Hummasti, S. R., Karry, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomcoman, M. A., Mazda, P., Moshfeghi, A. R., Moshireli, M., Nixon, K., Paciel, J. M., Park, S., Pfeiffer, B., Poorn, L., Punch, E., Sequelstra, A., Sethi, H., Sutr, E., Svrlckas, R. R., Tamey, B., Wan, K. H., Weinburg, T., Zhang, R., Zierian, L. L. and Rubin, G. M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Sequencing of Drosophila chromosome 2R, region 54E8-54F22
Unpublished (1998)
2 (bases 1 to 55452)
Celinker, S.E., George, R.A., Galle, R., Svirskaas, R.R., Hos

Agbayani, A., Arcalana, T. T., Baxter, E., Blazer, R. G., Chavez, C.,
Chev, M., Doyle, C. M., Farfan, D. E., Flanagan, J., Houston, K. A.,
Humstad, S. R., Karra, K., Kearney, L., Kim, S. H., Lee, B.,
Lomstad, M. A., Mak, J., Mazda, P., Mok, M. S., Mostrell, A. R.,
Mostrell, W., Nixon, K., Paceli, J. M., Park, S., Pfeiffer, B., PUNCH, E.,
Snir, E., Thomey, B., Wan, K. H., Whitelaw, K. R., Yee, A., Zhang, R.,
Zietan, L. L., and Kimmel, B. E.

COMMENT
On Nov 26, 1998 this sequence version replaced gi:3893057.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site ([http://www. fruitfly.berkeley.edu](http://www fruitfly.berkeley.edu)) or send email
to hdpe@fruitfly.berkeley.edu.

PI library location: 23-43.
Location/Qualifiers
1..53452

BASE COUNT	15556	a	12344	c	12189	g	15363	t
ORIGIN								

Query Match	55.5%;	Score 111;	DB 19;	Length 55452;
Best Local Similarity	41.7%;	Pred. No. 2.18e+01;		
Matches	20;	Conservative	10;	Mismatches 18;
			Indels	0;
			Gaps	0

```

Db 19408 TTTGGTTTTCGCACACTGGCCCCAAAAAATCCGCACGTAAAGTCTGGGAAT 19455
      ||::| :||| :|::| :|::| :|||
QY 1 TTYGGNGNNTTYACNGGNCMGNAARWSNCGMGNARRYTNGCNAAY 48

```

Search completed: Tue Apr 27 10:14:01 1999
Job time : 172 secs.

RESULT 2
ID W25219 standard; peptide: 17 AA.
AC W25219;
DE 12-JAN-1998 (first entry)
DE Rat orphanin FQ receptor-binding peptide inhibitor.
KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
KW morphine induced analgesia; methadone specific opioid receptor; ss.
OS Synthetic.
PN WO9707212-A1.
PD 27-FEB-1997.
PF 12-AUG-1996; U13305.
PR 03-NOV-1995; US-553058.
PR 11-AUG-1995; US-514451.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Bunzow JR, Civealli O, Grandy DK, Grisel JE, Mogil JS;
PI Monsma EJ, Notackner H-P, Reinscheid RK;
PI WPI: 97-165296/15.
DR Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
PT treatment of locomotor disease.
PS Claim 12; Page 48; 68pp; English.
CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (OFQ)
CC receptor.
CC Highly specific peptides that bind the OFQ receptor were identified as
CC OFQ receptor inhibitors. The peptides can be used to antagonise a
CC physiological effect of an opioid in an animal. The peptides antagonise
CC opiate induced hypothermia and morphine induced analgesia in animals.
CC They may also be used in the diagnosis and treatment of locomotor
CC disease. The peptides may also be used in the design of a methadone
CC specific opioid receptor (MSOR), in drug design and for the isolation
CC of endogenous receptors for anti-opioid agonists and antagonists found
CC in plasma, serum, lymph, cerebrospinal fluid etc.
SQ Sequence 17 AA;

Query Match 100.0%; Score 116; DB 25; Length 17;
Best Local Similarity 100.0%; Pred. No. 6.78e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 fggtgarkarklang 17
OY 1 FGFTGARKSARKLANQ 17

RESULT 3
ID W25220 standard; peptide: 17 AA.
AC W25220;
DE 12-JAN-1998 (first entry)
DE Rat orphanin FQ receptor-binding peptide inhibitor.
KW Orphanin FQ receptor; binding; locomotor disease; diagnosis;
KW treatment; opioid inhibitor; opiate induced hypothermia; drug design;
KW morphine induced analgesia; methadone specific opioid receptor; ss.
OS Synthetic.
PN WO9707212-A1.
PD 27-FEB-1997.
PF 12-AUG-1996; U13305.
PR 03-NOV-1995; US-553058.
PR 11-AUG-1995; US-514451.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Bunzow JR, Civealli O, Grandy DK, Grisel JE, Mogil JS;
PI Monsma EJ, Notackner H-P, Reinscheid RK;
PI WPI: 97-165296/15.
DR Mammalian orphanin FQ receptor inhibitor - used in the diagnosis and
PT treatment of locomotor disease.
PS Claim 12; Page 48; 68pp; English.
CC W25219 and W25220 are peptide inhibitors of the rat orphanin FQ (OFQ)
CC receptor.
CC Highly specific peptides that bind the OFQ receptor were identified as
CC OFQ receptor inhibitors. The peptides can be used to antagonise a
CC physiological effect of an opioid in an animal. The peptides antagonise
CC opiate induced hypothermia and morphine induced analgesia in animals.
CC They may also be used in the diagnosis and treatment of locomotor
CC disease. The peptides may also be used in the design of a methadone

CC specific opioid receptor (MSOR), in drug design and for the isolation
CC of endogenous receptors for anti-opioid agonists and antagonists found
CC in plasma, serum, lymph, cerebrospinal fluid etc.
SQ Sequence 17 AA;

Query Match 90.5%; Score 105; DB 25; Length 17;
Best Local Similarity 94.1%; Pred. No. 1.46e-04;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 fggtgarkarklang 17
OY 1 FGFTGARKSARKLANQ 17

RESULT 4
ID R38779 standard; peptide: 113 AA.
AC R38779;
DE 22-DEC-1993 (first entry)
DE Active site peptide fragment #18.
KW Active site; active region; prediction; secondary structure;
KW energy value; engineering.
OS Synthetic.
PN J05155899-A.
PD 22-JUN-1993.
PF 23-AUG-1991; 212284.
PR 23-AUG-1991; JP-212284.
PA (SAGA) SAGAMI CHEM RES CENTRE.
DR WPI: 93-232353/29.
PT Prediction of the active site in physiologically active
PT polypeptide - by determ. of sec. structure energy values of
PT partial regions along prim. sequence of physiologically active
PT polynucleotide, etc.
PS Disclosure; Page 24; 43pp; Japanese.
CC The sequences given in R38762-79 are peptides which were used to
CC demonstrate the method of the invention. These peptides represent
CC active sites/regions and the method of the invention may be used to
CC predict the active site of a polypeptide by determining the energy
CC values of the secondary structure of various partial regions of the
CC polypeptide to be tested and the primary sequence as well. At
CC least one partial region should show a local peak energy value
CC significantly higher or lower than the average standard energy value.
CC This method may be used in the elucidation of the mechanism of
CC action of polypeptides or polynucleotides, and in the engineered
CC improvement of such sequences.
SQ Sequence 113 AA;

Query Match 56.9%; Score 66; DB 8; Length 113;
Best Local Similarity 53.3%; Pred. No. 4.53e+00;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
Db 13 glsggrkpkrlvnr 27
OY 3 GFTGARKSARKLANQ 17

RESULT 5
ID R90989 standard; Protein: 529 AA.
AC R90989;
DE 27-MAY-1996 (first entry)
DE Human adrenergic G-protein coupled receptor.
KW G-protein coupled receptor polypeptide; GPCR; adrenergic receptor;
KW agonist; antagonist; therapy; hypertension; respiratory disease.
OS Homo sapiens.
PN WO9605225-A1.
PD 22-FEB-1996.
PF 10-AUG-1994; U09051.
PR 10-AUG-1994; WO-U09051.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Adams MD, Li Y, Soppet DR;
PI WPI: 96-139642/14.
DR N-PSDB: T13002.
PT New isolated G-protein coupled receptor polypeptide - used to
PT develop prods. to inhibit or stimulate adrenergic receptors for

PT treating e.g. hypertension or respiratory disorders
 PS Claim 4; Fig 1; 60pp; English.
 CC Human G-protein coupled receptor polypeptide (GCRP) (R90989)
 CC is a protein structurally related to the alpha-1 adrenergic
 CC receptor family. It can be produced by expression of a cDNA clone
 CC (T13002) isolated from a human infant brain cDNA library.
 CC Recombinant GCRP is used to screen for agonist and antagonist
 CC cpds. useful e.g. in the treatment of hypertension and respiratory
 CC disorders, and to raise anti-GCRP antibodies.
 SQ Sequence 529 AA;

Query Match 50.9%; Score 59; DB 16; Length 529;
 Best Local Similarity 50.0%; Pred. No. 2.53e+01;
 Matches 8; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 497 gfgfrrgsrlvsq 511
 |||:::|:
 QY 2 GFTGARKSARKLANQ 17

RESULT 6
 ID R76600 standard: Protein; 370 AA.
 AC R76600;
 DT 29-FEB-1996 (first entry)
 DE Chromobacterium esterase.
 KW Esterase; mass production; genetic engineering; plasmid; expression;
 KM recombinant.
 OS Chromobacterium sp.
 PN J07163364-A.
 PD 27-JUN-1995.
 PF 15-DEC-1993; 315497.
 PR 15-DEC-1993; JP-315497.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 95-260046/34.
 DR N-PSDB: Q92918.
 PT Esterase gene - for mass production of esterase
 PS Claim 1; Page 12-14; 15pp; Japanese.
 CC The chromobacterium esterase gene has been isolated and sequenced.
 CC It is found to be 1639 base pairs (bp) long with a CDS of 1110 bp
 CC which codes for the 370 amino acid esterase protein. The gene
 CC can be used in recombinant systems to allow the mass prodn. of
 CC esterase in high yields.
 SQ Sequence 370 AA;

Query Match 49.1%; Score 57; DB 14; Length 370;
 Best Local Similarity 33.3%; Pred. No. 4.09e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 138 gfgfrrgsrlvsq 152
 |||:::|:
 QY 3 GFTGARKSARKLANQ 17

RESULT 7
 ID R86303 standard: protein; 370 AA.
 AC R86303;
 DT 16-APR-1996 (first entry)
 DE Thermostable esterase with mutation at position 160 and/or 189.
 KW Thermostable; esterase; mutant; organic synthesis; site specific.
 OS Chromobacterium sp.
 FH Key Location/Qualifiers
 FT misc-difference 160
 FT /note- "site specific mutation, optional when
 FT wild type Gly189 is also mutated"
 FT misc-difference 189
 FT /note- "site specific mutation, optional when
 FT wild type Gly160 is also mutated"
 PN J07213280-A.
 PD 15-AUG-1995.
 PF 02-FEB-1994; 011014.
 PR 02-FEB-1994; JP-011014.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 95-315922/41.

PT Thermostable esterase with mutation at amino acid position 160
 PT and/or 189 - useful for organic synthesis, also mutant gene and
 PT transformed microbe for recombinant prodn. of enzyme
 PS Claim 1; Pages 8-9; 16pp; Japanese.
 CC A thermostable Chromobacterium sp. esterase is produced when the
 CC wild type amino acid residues Gly160 and/or Gly189 undergo site
 CC specific mutation(s). The mutant thermostable esterase (R86303)
 CC can be used in organic synthesis.
 SQ Sequence 370 AA;

Query Match 49.1%; Score 57; DB 15; Length 370;
 Best Local Similarity 33.3%; Pred. No. 4.09e+01;
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 138 gfgfrrgsrlvsq 152
 |||:::|:
 QY 3 GFTGARKSARKLANQ 17

RESULT 8
 ID R29516 standard: Protein; 1385 AA.
 AC R29516;
 DT 20-APR-1993 (first entry)
 DE Bt toxin 17a.
 KW nematode worms; nematicide; nematicidal toxin; agriculture; plants;
 KM crops; pests; CryV proteins.
 OS Bacillus thuringiensis.
 PN EP-517367-A.
 PD 09-DEC-1992.
 PF 01-MAY-1992; 303969.
 PR 03-MAY-1991; US-693018.
 PR 31-JAN-1992; US-830050.
 PR 23-APR-1992; US-871510.
 PA (MYCO) MYCOGEN CORP.
 PI Fonceriada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI: 92-408829/50.
 DR N-PSDB: Q32159.
 PT Nematocidal toxins from Bacillus thuringiensis - useful for
 PT control of animal or plant parasites, deoxyribonucleic acid
 PT coding sequences, transformed hosts and transgenic plants
 PS Claim 1(b); Page 23; 57pp; English.
 CC This sequence represents the Bacillus thuringiensis delta-endotoxin
 CC 17a having nematicidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematicidal
 CC compans, either sprayed on plants to kill agricultural crop pests,
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgekkalkrlvng 956
 |||:::|:
 QY 4 FTGARKSARKLANQ 17

RESULT 9
 ID R28889 standard: Protein; 1385 AA.
 AC R28889;
 DT 01-APR-1993 (first entry)
 DE Toxin 17a.
 KW Endotoxin; acarides; pest; Two Spotted spider; mite;
 KW phytophagous.
 OS Bacillus thuringiensis PS17.
 FH Key Location/Qualifiers
 FT peptide 308..315
 FT /note- "amino acid sequence of a probe for cloning
 FT acaride-active genes"
 PN M09219106-A.
 PD 12-NOV-1992.
 PF 30-APR-1992; U03546.
 PR 30-APR-1991; US-6932210.

PR 13-SEP-1991: US-759248.
 PR 30-SEP-1991: US-768141.
 PA (MYCO) MYCOGEN CORP.
 PI Bagley AL, Cannon RJC, Payne JM;
 DR WPI: 92-398411/48.
 DR N-PSDB: Q20803.
 PT New Bacillus thuringiensis isolates and toxins - used for
 PT controlling acarid pests of livestock, fowl, stored prods. and
 PT plants
 PS Claim 18: Page 53 + 27-30: 62pp; English.
 CC Gene sequences encoding a toxin which is active against acarides and
 CC is obtainable from B. thuringiensis isolates PS17a, PS17b, 33f2,
 CC PS52a1, PS59D1, PS66a1 and PS50C are given in Q30803-07 and Q30820-21
 CC respectively. The toxin is a delta-endotoxin active against acarid
 CC pests, including the Two Spotted Spider mite. The isolates can be
 CC used against non-phytophagous mites such as acarid pests of livestock,
 CC fowl and stored prods. The genes can be cloned and used to
 CC transform other hosts, which can be used to control mites, or in
 CC the case of transgenic plants, be resistant to mites.
 SQ Sequence 1385 AA;
 SO
 Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 943 fgdekalkrlvng 956
 | : | | | |
 | : | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 10
 ID R58631 standard; Protein: 1385 AA.
 AC R58631;
 DE 16-MAY-1995 (first entry)
 DE Bacillus thuringiensis PS17a delta-endotoxin protein.
 DE Endotoxin; Bacillus thuringiensis; PS17; toxin; delta-endotoxin;
 KW parasporal; crystalline; inclusion; toxic; pest; spider mite;
 KW Tetranychus urticae; house dust mite; Dermatophagoides pteromyssinus;
 KW phytophagous; mite; acarid; livestock; fowl; acaridicidal; insect;
 KW virus; plant; control; transgenic; resistance.
 OS Bacillus thuringiensis strain PS17.
 PS US5350576-A.
 PD 27-SEP-1994.
 PF 13-SEP-1991: 759248.
 PR 13-SEP-1991: US-759248.
 PR 30-SEP-1991: US-768141.
 PR 17-MAY-1993: US-063170.
 PA (MYCO) MYCOGEN CORP.
 PI Cannon RJC, Payne JM, Ralph AL;
 DR WPI: 94-310547/38.
 DR N-PSDB: Q72179.
 PT Controlling house dust mite - using new delta endotoxin isolates
 PT from Bacillus thuringiensis with acaridicidal activity
 PS Claim 5: Column 13-16: 10pp; English.
 CC The amino acid sequence of a novel delta-type endotoxin from Bacillus
 CC thuringiensis (B.t.) strain PS17. The toxin gene encodes a protein of
 CC 1385 amino acids. The PS17 delta-endotoxins range in mol. wt. from 128
 CC to 155 kDa. The delta-endotoxins are derived from parasporal crystalline
 CC protein inclusions and are highly toxic to pests and specific in their
 CC activity. The toxins presented (Q72179-80) are specific to the
 CC two-spotted spider mite (Tetranychus urticae) and to the house dust mite
 CC (Dermatophagoides pteromyssinus). The use of B.t. isolates containing
 CC the endotoxins can also target non-phytophagous mites such as acarid
 CC pests of livestock, fowl and stored products. The gene(s) from the B.t.
 CC isolates which encode the acaridicidal endotoxin can be cloned from the
 CC isolates and used to transform other hosts e.g. prokaryotic cells, insect
 CC viruses, eukaryotic cells or plants, which could be used to control mites
 CC or, in the case of transgenic plants, confer resistance to mites.
 SQ Sequence 1385 AA;
 SO
 Query Match 48.3%; Score 56; DB 12; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 943 fgdekalkrlvng 956
 | : | | | |
 | : | | | |
 QY 4 FTGARKSARKLANQ 17

Db 943 fgdekalkrlvng 956
 | : | | | |
 | : | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 11
 ID W13884 standard; Protein: 1385 AA.
 AC W13884;
 DE 14-MAY-1997 (first entry)
 DE 17a toxin.
 KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;
 KW biological control; Monomorium pharaonis; delta-endotoxin; lepidoptera;
 KW insect.
 OS Bacillus thuringiensis isolate PS17a.
 PS US5596071-A.
 PD 21-JAN-1997.
 PF 22-MAY-1991: 703977.
 PR 22-MAY-1991: US-703977.
 PR 25-NOV-1991: US-797645.
 PR 22-MAY-1992: US-887980.
 PR 24-NOV-1993: US-158232.
 PA (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Fu J, Kennedy MK, Meier H, Payne JM;
 PI Randall JB, Schnepf HE, Schwab GE, Vick HJ;
 DR WPI: 97-107615/10.
 DR N-PSDB: T60069.
 PT Bacillus thuringiensis toxin - active against hymenopteran pests
 PS Disclosure: Column 35-42: 64pp; English.
 CC This sequence represents the 17a toxin isolated from the Bacillus
 CC thuringiensis (B.t.) isolate PS17a. B.t. is a gram-positive, spore
 CC forming, soil bacterium, characterised by parasporal crystalline protein
 CC inclusions. These proteins can be highly toxic to pests, and have been
 CC used to produce insect resistant plants. The previously isolated B.t.
 CC delta-endotoxins were mainly active against lepidopteran insects, however
 CC the proteins of the invention are active against hymenopteran insects.
 CC This sequence is an example of a toxin of the invention, for which the
 CC sequences shown in W13888 and W13871 represent the generic formulae. As
 CC the toxins of the invention are active against hymenopteran pests, they
 CC can be used for the biological control of ants, particularly pharaoh ants
 CC (Monomorium pharaonis).
 SQ Sequence 1385 AA;
 SO
 Query Match 48.3%; Score 56; DB 21; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 943 fgdekalkrlvng 956
 | : | | | |
 | : | | | |
 QY 4 FTGARKSARKLANQ 17
 RESULT 12
 ID R28803 standard; Protein: 1385 AA.
 AC R28803;
 DE 30-MAR-1993 (first entry)
 DE Bt toxin 17a.
 KW nematode worms; nematocidal; nematocidal toxin; agriculture; plants;
 KW crops; pests; Cyy proteins.
 OS Bacillus Thuringiensis.
 PS W09219739-A.
 PD 12-NOV-1992.
 PF 01-MAY-1992: U03624.
 PR 03-MAY-1991: US-693018.
 PR 31-JAN-1992: US-830050.
 PR 23-APR-1992: US-871510.
 PA (MYCO) MYCOGEN CORP.
 PI Foncerrada L, Narva KE, Payne JM, Schnepf HE, Schwab GE;
 DR WPI: 92-39866/48.
 DR N-PSDB: Q30935.
 PT New genes and toxins against nematodes - obtd. from Bacillus
 PT Thuringiensis isolates with nematocidal activity
 PS Claim 1(b); Page 42; 77pp; English.

CC This sequence represents the *Bacillus thuringiensis* delta-endotoxin
 CC 17a having nematocidal activity. It was decoded from the appropriate
 CC DNA. Intact cells expressing this toxin may be used in nematocidal
 CC compounds, either sprayed on plants to kill agricultural crop pests,
 CC or taken as a drug to kill gastro-intestinal worms in mammalian hosts.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrklyng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 13
 ID R29026 standard; Protein; 1385 AA.
 AC R29026;
 DT 20-APR-1993 (first entry)
 DE *Bacillus thuringiensis* toxin protein 17a.
 KW Toxin protein; ant.
 OS *Bacillus thuringiensis* strain PS17.
 PN MO9220802-A.
 PD 26-NOV-1992.
 PF 22-MAY-1992; U04316.
 PR 22-MAY-1991; US-703977.
 PR 25-NOV-1991; US-797645.
 PR 12-MAY-1992; EP-304228.
 PA (MYCO) MYCOGEN CORP.
 PI Kennedy MK, Meier H, Payne JM, Randall JB, Uick HJ;
 DR WPI; 92-415780/50.
 DR N-PSDB; 031409.
 PT Toxin proteins isolated from *Bacillus thuringiensis* - for controlling
 PT ants, e.g. fire, carpenter, Argentine and pharaoh ants
 PS Disclosure: Page 35; 71pp; English.
 CC *Bacillus thuringiensis* toxin protein 17a is useful as a method of
 CC biological control of ants, e.g. fire ants, carpenter ants,
 CC Argentine ants and pharaoh ants, as an alternative to chemical
 CC insecticides.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 6; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrklyng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 14
 ID R76112 standard; Protein; 1385 AA.
 AC R76112;
 DT 27-FEB-1996 (first entry)
 DE *PS17a* acaricide-active toxin from strain PS17.
 KW Insecticide; toxin; acaricide-active toxin; delta-endotoxin; PS17a; PS17;
 OS two spotted spider mite; tetranychus urticae; *Bacillus thuringiensis*.
 KM *Bacillus thuringiensis* strain PS17.
 FT Key location/Qualifiers
 FT region 2.15
 FT US5424410-A. /note- "N-terminal region"
 PN 13-JUN-1995.
 PF 30-APR-1991; 693210.
 PR 30-APR-1991; US-693210.
 PR 13-SEP-1991; US-759248.
 PR 30-SEP-1991; US-768141.
 PR 30-APR-1992; US-876280.
 PR 03-NOV-1993; US-147188.
 PA (MYCO) MYCOGEN CORP.
 PI Bagley AL, Cannon RJC, Payne JM;
 DR WPI; 95-223727/29.

DR N-PSDB; 094053.
 PT DNA encoding delta endotoxin of *Bacillus thuringiensis* - used for
 PT the control of acaride pests and for the production of acaride pest
 PT resistant plants
 PS Example 3; Column 27-34; 42pp; English.
 CC The sequences represented by R76112 and R76113 represent acaride-active
 CC toxins isolated from *B. thuringiensis* isolate PS17. This sequence
 CC represents PS17a. These sequences are delta-endotoxins.
 CC Delta-endotoxins are insect toxins which act on insects in the orders
 CC Lepidoptera and Diptera. The toxin acts to destroy insect gut epithelium
 CC cells. The toxin is ingested in crystalline form, by larvae, and is
 CC transformed into biologically active forms by the insect gut juice
 CC proteases (in susceptible insects). The N-terminal region of this
 CC sequence was used in the design of probes for this sequence and other
 CC acaride-active toxins. The DNA encoding this sequence can be inserted
 CC into a vector which can be introduced into plant cells, thereby
 CC conferring insect resistance. This protein can be combined with other
 CC insecticidal proteins to increase the spectrum of activity and to give
 CC complete control of target pests. Plants transformed with the DNA
 CC encoding this sequence become resistant to acaride pests, e.g. the two
 CC spotted spider mite, *Tetranychus urticae*.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 14; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrklyng 956
 | : | | | | |
 QY 4 FTGARKSARKLANQ 17

RESULT 15
 ID R20066 standard; Protein; 1385 AA.
 AC R20066;
 DT 26-MAR-1992 (first entry)
 DE *B. thuringiensis* toxin PS17a.
 KW worm; nematocidal; flukicide; anthelmintic; parasite.
 OS *Bacillus thuringiensis*.
 PN EP-462721-A.
 PD 27-DEC-1991.
 PF 04-JUN-1991; 305047.
 PR 11-JUN-1990; US-535810.
 PR 24-JUL-1990; US-557246.
 PR 27-JUL-1990; US-558738.
 PR 10-AUG-1990; US-565544.
 PR 14-MAR-1991; US-669126.
 PR 27-MAR-1991; US-675772.
 PR 03-MAR-1991; US-695018.
 PA (MYCO-) MYCOGEN CORP.
 PI Narva KE, Payne JM, Schwab GE, Hickie LA, Galasan T;
 PI Sick AJ;
 DR WPI; 92-001086/01.
 DR N-PSDB; 020334.
 PT New *Bacillus thuringiensis* strains expressing toxins - have
 PT nematocidal activity, to control nematodes, helminths and flukes
 PT e.g. liver fluke *Fasciola hepatica*
 PS Claim 6; Page 24; 47pp; English.
 CC The PS17a toxin gene was isolated from B.t. strain PS17. Probe
 CC Q20338 was used to screen a PS17 total cellular DNA library. The
 CC gene was subcloned in a vector to give plasmid pMYC331 which was
 CC transformed into the acrycristiferous B.t. strain, HD-1 cryB. The
 CC recombinant B.t. strain was grown to sporulation and the toxin
 CC protein purified by NaBr gradient centrifugation.
 CC See also Q20335-Q20343.
 SQ Sequence 1385 AA;

Query Match 48.3%; Score 56; DB 3; Length 1385;
 Best Local Similarity 50.0%; Pred. No. 5.19e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 fgdekaklrklyng 956
 | : | | | | |

Wed Apr 28 09:08:31 1999

US-09-011-797-2.rag

Page 6

OY 4 FTGARKSARKLANQ 17

Search completed: Fri Apr 16 14:17:41 1999
Job time : 17 secs.

 M P E R E L I
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mperch_Pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 16 14:19:10 1999; MasPar time 4.96 Seconds

Tabular output not generated. 170.573 Million cell updates/sec

Title: >US-09-011-797-2

Description: (1-17) from US09011797.pep

Perfect Score: 116

Sequence: 1 FGFTGARKSARKLANQ 17

Scoring table:

PAM 150
 Gap 15

Searched: 165420 segs, 49795644 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl6
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 26.722; Variance 32.797; scale 0.815

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	176	6	062647	8.09e-15
2	66	56.9	526	1	P95887	4.09e-02
3	60	51.7	776	4	060719	7.95e-01
4	59	50.9	177	1	028463	1.28e+00
5	59	50.9	211	2	066484	1.28e+00
6	57	49.1	133	4	098989	3.27e+00
7	57	49.1	152	10	041349	3.27e+00
8	57	49.1	614	4	043245	3.27e+00
9	57	49.1	801	4	013436	3.27e+00
10	57	49.1	1061	10	004987	3.27e+00
11	56	48.3	890	10	P93024	5.18e+00
12	56	48.3	902	10	064965	5.18e+00
13	56	48.3	1385	2	045760	5.18e+00
14	55	47.4	470	10	040346	8.16e+00
15	55	47.4	476	10	040347	8.16e+00
16	54	46.6	224	1	027716	1.28e+01
17	54	46.6	316	2	033872	1.28e+01
18	54	46.6	378	5	016903	1.28e+01
19	54	46.6	593	10	P93002	1.28e+01
20	54	46.6	651	2	069635	1.28e+01

Result ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	116	100.0	176	6	062647	8.09e-15
2	66	56.9	526	1	P95887	4.09e-02
3	60	51.7	776	4	060719	7.95e-01
4	59	50.9	177	1	028463	1.28e+00
5	59	50.9	211	2	066484	1.28e+00
6	57	49.1	133	4	098989	3.27e+00
7	57	49.1	152	10	041349	3.27e+00
8	57	49.1	614	4	043245	3.27e+00
9	57	49.1	801	4	013436	3.27e+00
10	57	49.1	1061	10	004987	3.27e+00
11	56	48.3	890	10	P93024	5.18e+00
12	56	48.3	902	10	064965	5.18e+00
13	56	48.3	1385	2	045760	5.18e+00
14	55	47.4	470	10	040346	8.16e+00
15	55	47.4	476	10	040347	8.16e+00
16	54	46.6	224	1	027716	1.28e+01
17	54	46.6	316	2	033872	1.28e+01
18	54	46.6	378	5	016903	1.28e+01
19	54	46.6	593	10	P93002	1.28e+01
20	54	46.6	651	2	069635	1.28e+01

DR EMBL: Y08256; E283902; -
SQ SEQUENCE 526 AA; 56210 MW; 5AD5F9A0 CRC32;

Query Match 56.9%; Score 66; DB 1; Length 526;
Best Local Similarity 52.9%; Pred. No. 4.09e-02;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 220 FGGFGAFHNAQTSSQ 236
1 FGGFTGARKSARKLANQ 17

RESULT 3
ID 060719 PRELIMINARY; PRT; 776 AA.
AC 060719:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CULLIN 1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CERVIX;
RA MICHEL J.J., XIONG Y.;
RL CELL GROWTH DIFFER. 0:0-0(1998).
DR EMBL: AF062536; G3139077;
DR PROSITE: PS01256; CULLIN_1;
SQ SEQUENCE 776 AA; 89678 MW; 480E807E CRC32;

Query Match 51.7%; Score 60; DB 4; Length 776;
Best Local Similarity 46.7%; Pred. No. 7.95e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Db 566 FTATYASRHSGRKLT 580
1 FGGFTGARKSARKLANQ 15

RESULT 4
ID 028463 PRELIMINARY; PRT; 177 AA.
AC 028463:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 20.6 KD PROTEIN.
GN AF1812
OS ARCHAEOGLOBUS FULGIDUS.
OC ARCHAEBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
RA FLEISCHMANN R.D., QUTACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBERCK T.,
RA COTTON M.D., SPRIGGS T.D., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., MOESE C.R.,
RA VENTER J.C.;
RL NATURE 390:364-370(1997).
DR EMBL: AE000978; G2648748;
DR TIGR: AF1812;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 177 AA; 20596 MW; 9721D71F CRC32;

Query Match 50.9%; Score 59; DB 1; Length 177;
Best Local Similarity 46.7%; Pred. No. 1.28e+00;

Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 138 FLAGSGIRKARLKG 152
1 FGGFTGARKSARKLANQ 15

RESULT 5
ID 066484 PRELIMINARY; PRT; 211 AA.
AC 066484:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE RIBOSOMAL PROTEIN S04.
GN RPSD.
OS ACUIPEX AEOLICUS.
OC EUBACTERIA; ACUIFICAELES; ACUIFICACEAE; ACUIPEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUIJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUIJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000672; G2982819;
DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 211 AA; 24615 MW; 8A762049 CRC32;

Query Match 50.9%; Score 59; DB 2; Length 211;
Best Local Similarity 40.0%; Pred. No. 1.28e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 108 GPASTRQARQOLVAH 122
3 GFTGARKSARKLANQ 17

RESULT 6
ID 099899 PRELIMINARY; PRT; 133 AA.
AC 099899:
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE P65-MEMBER OF THE STEROID/THYROID RECEPTOR SUPERFAMILY (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96253399.
RA HANAUSEK M., SZEMRAJ J., ADAMS A.K., WALASZEK Z.;
RL CANCER DETECT. PREV. 20:94-102(1996).
DR EMBL: S82307; G1839614;
DR PFWA: PF00105; zf-C4
FT NON_TER 133
SQ SEQUENCE 133 AA; 15100 MW; B7F06493 CRC32;

Query Match 49.1%; Score 57; DB 4; Length 133;
Best Local Similarity 53.8%; Pred. No. 3.27e+00;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 28 GGRAGPRRGARRL 40
2 GFTGARKSARKLANQ 14


```

FT NON_TER 1 1
SQ SEQUENCE 890 AA; 98323 MW; DC5A2F01 CRC32;
Query Match 48.3%; Score 56; DB 10; Length 890;
Best Local Similarity 53.8%; Pred. No. 5.18e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 159 GGFVSPRRARL 171
   111:111:11
QY 2 GGFGARSRARL 14

RESULT 12
ID 064965; PRELIMINARY; PRT; 902 AA.
AC 064965;
DR 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR.
GN MP.
OS ABABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAPRARIACE; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA.
RA HARDTKE C.S., BERLETH T.;
RL EMBO J. 17:1405-1411(1998).
DR EMBL; AF037229; G2982222; -.
SQ SEQUENCE 902 AA; 99662 MW; 667EBA47 CRC32;

Query Match 48.3%; Score 56; DB 10; Length 902;
Best Local Similarity 53.8%; Pred. No. 5.18e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 171 GGFVSPRRARL 183
   111:111:11
QY 2 GGFGARSRARL 14

RESULT 13
ID 045760; PRELIMINARY; PRT; 1385 AA.
AC 045760;
DR 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE DELTA-ENDOTOXIN (FRAGMENT).
GN CRVA.
OS BACILLUS THURINGIENSIS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PS17;
RA NARVA K.E., PAYNE J.M., SCHWAB G.E., HICKLE L.A., GALASAN T.,
RA STICK A.J.;
RL SUBMITTED (DEC-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L07025; G142869; -.
DR PFAM; PF00555; endotoxin.
FT NON_TER 1385 1385
SQ SEQUENCE 1385 AA; 152439 MW; 2592A8C7 CRC32;

Query Match 48.3%; Score 56; DB 2; Length 1385;
Best Local Similarity 50.0%; Pred. No. 5.18e+00;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 943 FGDEKALRLVNO 956
   111:111:11
QY 4 FTGARSRARLNO 17

RESULT 14
ID 040346; PRELIMINARY; PRT; 470 AA.
AC 040346;

```

```

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE LEGUMIN PRECUR PRECURSOR (FRAGMENT).
OS MAGNOLIA SALICIFOLIA
OC EUKARYOTA; PLANTAE; EMBRYOIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC MAGNOLIACEAE; MAGNOLIALES; MAGNOLIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SINGLE TREE IN MUNICH BOTANICAL GARDEN; TISSUE-SEED;
RX MEDLINE; 95278215.
RA FISCHER H., HAAKE V., HORSTMANN C., JENSEN U.;
RL EUR. J. BIOCHEM. 229:645-650(1995).
DR EMBL; X82463; G793854; -.
DR PFAM; PF00190; Seedstore_11s.
KW SIGNAL.
FT NON_TER 1 1
FT SIGNAL <1 19 POTENTIAL.
SQ SEQUENCE 470 AA; 52986 MW; EEB11097 CRC32;

Query Match 47.4%; Score 55; DB 10; Length 470;
Best Local Similarity 46.7%; Pred. No. 8.16e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 236 AFGVRETRARLQSO 250
   111:111:11
QY 3 GFTGARSRARLNO 17

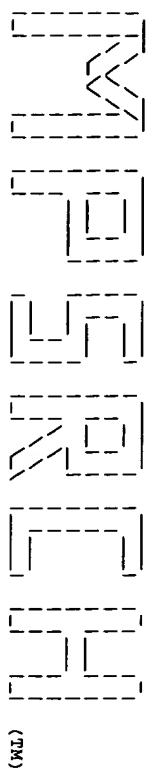
RESULT 15
ID 040347; PRELIMINARY; PRT; 476 AA.
AC 040347;
DR 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE GLOBULIN PRECURSOR.
OS MAGNOLIA SALICIFOLIA.
OC EUKARYOTA; PLANTAE; EMBRYOIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;
OC MAGNOLIACEAE; MAGNOLIALES; MAGNOLIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SINGLE TREE IN MUNICH BOTANICAL GARDEN; TISSUE-SEED;
RX MEDLINE; 95278215.
RA FISCHER H., HAAKE V., HORSTMANN C., JENSEN U.;
RL EUR. J. BIOCHEM. 229:645-650(1995).
DR EMBL; X82464; G793856; -.
DR PFAM; PF00190; Seedstore_11s.
KW SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
SQ SEQUENCE 476 AA; 53663 MW; 5435E3FC CRC32;

Query Match 47.4%; Score 55; DB 10; Length 476;
Best Local Similarity 46.7%; Pred. No. 8.16e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 240 AFGVRETRARLQSO 254
   111:111:11
QY 3 GFTGARSRARLNO 17

Search completed: Fri Apr 16 14:19:42 1999
Job time : 32 secs.

```



Release 3.1a John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_tpr n.a. n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:17:01 1999; Maspar time 25.49 Seconds

Tabular output not generated. 272.182 Million cell updates/sec

Title: >US-09-011-797-2
Description: (1-17) from US09011797.pep
Perfect Score: 200
N.A. Sequence: 1 TTYGNGNGNTTACNGNGC.....CNGMNAHYTNGCAAYCAR 51
Comp: AARCCNCNARFGNCNCG.....GNKNCNTTYRANGCNRTRGV

Scoring table: TABLE bkttranslated
Gap 40

Nmatch STD : Dbase 0: Query 0

Searched: 188442 segs, 68026449 bases x 2

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

n:geneg32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 34.207; Variance 134.692; scale 0.254

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	196	98.0	932	34	T79888	DNA encoding human op
2	106	53.0	1656	33	T85271	HMG-CoA reductase deg
3	96	48.0	3411	40	V16518	DNA encoding a Bacill
4	96	48.0	580073	27	T58840	Mycoplasma genitalium
5	95	47.5	359	7	O46579	Active site polynucle
6	91	45.5	297	8	O60810	Human brain Expressed
7	91	45.5	1302	38	T90224	Magnetostrictulum sp.
8	91	45.5	1606	3	O20264	Mec-4 gene.
9	91	45.5	1607	6	O38934	mec-4 gene.
10	91	45.5	3471	40	V16515	DNA encoding a Bacill
11	91	45.5	36335	16	O68003	Ad2/-ORF6/PGK-CFR nu
12	90	45.0	3872	17	T10156	Sockeye salmon growth

RESULT	ID	Score	Query Match	Length	DB ID	Description	Pred. No.
13	90	45.0	4086	9	O51033	Human myod gene and m	
14	90	45.0	4281	9	O53471	elk cDNA.	
15	90	45.0	24417	38	T97221	Pseudomonas aeruginos	
16	90	45.0	134525	1	O04525	Total base sequence o	
17	90	45.0	134525	1	O04525	Total base sequence o	
18	86	43.0	364	20	T16817	Kaposi's sarcoma asso	
19	86	43.0	439	19	T20045	Human gene signature	
20	86	43.0	447	1	N90638	Aspergillus nidulans	
21	86	43.0	594	24	T33926	Sequence encoding pol	
22	86	43.0	1215	2	N60836	5' region of bacterial	
23	86	43.0	1535	23	T33656	M. tuberculosis cellu	
24	86	43.0	1738	23	T40760	Dcc2 (brain-specific	
25	86	43.0	1738	10	O62817	Genomic sequence enco	
26	86	43.0	1738	16	T03128	Alpha-1B adrenergic r	
27	86	43.0	1738	10	O63181	Alpha 1b adrenergic r	
28	86	43.0	1914	18	T09274	MCF FKN provirus env	
29	86	43.0	2233	5	O30454	Sequence of ACP (acyl	
30	86	43.0	2856	34	T69800	Autobasidin sensiti	
31	86	43.0	2940	31	T63334	Human host cell prote	
32	86	43.0	3075	37	O06252	Murine putromycin-sen	
33	86	43.0	3108	8	O46248	Aspergillus niger cat	
34	86	43.0	3108	8	O46248	Human putromycin-sensi	
35	86	43.0	3153	37	O06251	Human putromycin-sensi	
36	86	43.0	3362	37	O06278	Human putromycin-sensi	
37	86	43.0	3436	11	O68691	Human endothelin-2 ge	
38	86	43.0	3747	37	O06250	Human putromycin-sensi	
39	86	43.0	4140	34	T69799	Autobasidin resistanc	
40	86	43.0	4414	1	N92771	HIV-2 variant HIV-D20	
41	86	43.0	5309	2	O10280	Complete sequence of	
42	86	43.0	8341	28	T09700	DNA encoding genes in	
43	86	43.0	8528	8	O46249	Construct ECL2 (Conla	
44	86	43.0	20710	21	T30681	Kaposi's sarcoma asso	
45	86	43.0	20710	20	T16806	Kaposi's sarcoma asso	

ALIGNMENTS

RESULT 1
ID T79888 standard: DNA; 932 BP.
AC T79888;
DT 08-DEC-1997 (first entry)
DE DNA encoding human opiod receptor-like 1 receptor ligand.
KW Opiod receptor-like 1; ORL1; ligand; pronociceptive; stress;
KW hyperalgesia; locomotor activity disease; anxiety; neuroendocrine;
KW memory; attention; sensory perception; learning; homeostasis;
KW hypoaesthesia; nociceptin; endorphin; dynorphin A; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT cds 1..507
FT /tag a
FT /note "no start codon"
PN M09707208-A1.
PD 27-FEB-1997.
PF 14-AUG-1996; BE0087.
PR 15-AUG-1995; US-002368.
PA (ULBR) UNIT LIBRE BRUXELLES.
PI Meunier J, Mollereau C, Parmentier M, Vassart G;
DR P-PSDB: W25162, W25163, W25164.
DR Novel ligand for the opiod receptor-like receptor, nociceptin -
PT has pro-nociceptive properties, useful for treating or preventing
PT diseases related to e.g. stress, hyperalgesia, locomotor activity,
PT etc.
PT claim 1: Page 1: 48pp; English.
CC T79888 encodes a ligand to human opiod receptor-like 1 (ORL1) receptor
CC designated nociceptin, which resembles the endorphin dynorphin A.
CC Inhibitors of nociceptin or its DNA, e.g. antisense sequences and
CC antibodies can be used as new types of drugs in the control of various
CC behaviours or functions. The inhibitors can be used to prevent or treat
CC a disease related to hyperalgesia, neuroendocrine secretion, stress,
CC locomotor activity, anxiety, instinctive behaviour, learning and memory,
CC homeostasis, hyperalgesia, hypoaesthesia and/or sensory perception.
SQ Sequence 932 BP; 224 A; 273 C; 238 G; 197 T;

```

Query Match          98.0%   Score 196; DB 34-DB; Length 932;
Best Local Similarity 56.9%; Pred. No. 5,72e+08;
Matches 29; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

Db      364 ttgggggcttcactggcgccgcgaatcgaagccgccgaagtgtgcacaccg 414
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      1 tTtGNGNCNTTYACNGCNGCMNNAARMSNCCNMGNARYTCNNCAAYCAR 51

RESULT      2
ID      T85271 standard; DNA; 1656 BP.
AC      T85271;
DT      08-DEC-1997 (first entry)
DE      HMG-CoA reductase degradation polypeptide 1 HRD1 gene.
KW      3-hydroxy-3-methylglutaryl: coenzyme A; cholesterol; Hrd1p; Hrd2p;
OS      Saccharomyces cerevisiae.
FH      Key Location/Qualifiers
FT      cds 1..1656
           /*tag= a
           /product= Hrd1p

FN      W09707219-A2.
PD      27-FEB-1997.
PE      16-AUG-1996; IB1161.
PR      17-AUG-1995; US-002381.
PA      (REGC ) UNIV CALIFORNIA.
PI      Hampton R. Rine JD;
DR      WPI. 97-165303/15.
OR      P-P5DB; W27150.
PT      3-Hydroxy-3-methyl:glutaryl CoA reductase degradation polypeptide(s)
PS      - useful as therapeutic agents to reduce hypercholesterolemia
SC      Claim 2; Page 88-89; 132pp; English.
CC      The present sequence represents the 3-hydroxy-3-methylglutaryl
CC      (HMG)-CoA reductase degradation (HRD) protein HRD1 gene. The HRD1 gene
CC      encodes the Hrd1 protein. Hrd proteins can be used to regulate the
CC      degradation of HMG-CoA reductase, e.g., as therapeutic agents to reduce
CC      hypercholesterolemia, and to elucidate how the cholesterol pathway
CC      modulates the degradation of HMG-CoA reductase. In addition, as a result,
CC      of their ability to bind the proteasome complex, antibodies that
CC      specifically bind hrd polypeptides can be used to isolate the proteasome
CC      complex. Further, they can be used in various assays to identify
CC      compounds that modify the degradation of HMG-CoA reductase independently
CC      of the beneficial IDL receptor control axis. The nucleic acid molecules
CC      can be used as molecular probes for the isolation of homologous nucleic
CC      acid molecules and for the detection of HRD nucleic acid molecules in
CC      yeast.
SQ      Sequence 1656 BP; 537 A; 324 C; 304 G; 491 T;

Query Match          53.0%   Score 106; DB 33; Length 1656;
Best Local Similarity 46.2%; Pred. No. 5,00e+00;
Matches 18; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Db      1555 acgggggcgagcaaatgccagaanaattgtcatcacccg 1633
        ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy      13 ACNGNGCNGCMNNAARMSNCCNMGNARYTCNNCAAYCAR 51

RESULT      3
ID      V16518 standard; DNA; 3411 BP.
AC      V16518;
DT      11-JUN-1998 (first entry)
DE      DNA encoding a Bacillus thuringiensis toxin.
KW      Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;
OS      Bacillus thuringiensis.
FH      Key Location/Qualifiers
FT      cds 1..3405
           /*tag= a

FN      W09800546-A2.
PD      08-JAN-1998.
PE      01-JUL-1997; U11658.
PR      01-JUL-1996; US-674002.

```

[illegible]

```
FT      cds      hydrofolate dehydrogenase (fold) from E. coli"
FT      17474..19243
FT      /*tag- f
FT      /label- MG015
FT      /note- "Previously identified as MORF-20084, the
FT      encoded protein shows 32.23 percentage
FT      identity to transport ATP-binding protein
FT      (meba) from E. coli"
FT      26478..27344
FT      cds      /*tag- g
FT      /label- MG023
FT      /note- "Previously identified as MORF-20092, the
FT      encoded protein shows 45.96 percentage
FT      identity to fructose-bisphosphate aldolase
FT      (far) from B. subtilis"
FT      27345..28448
FT      cds      /*tag- h
FT      /label- MG024
FT      /note- "Previously identified as MORF-19826 and
FT      MORF-20093, the encoded protein shows 46.84
FT      percentage identity to GTP-binding protein
FT      from E. coli"
FT      36987..38978
FT      cds      /*tag- i
FT      /label- MG032
FT      /note- "Previously identified as MORF-20099, the
FT      encoded protein shows 26.82 percentage
FT      identity to ATP-dependent nuclease (adda)
FT      from B. subtilis"
FT      39242..39904
FT      cds      /*tag- j
FT      /label- MG033
FT      /note- "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (gufp) from B. subtilis"
FT      complement (39873..40514)
FT      /*tag- k
FT      /label- MG034
FT      /note- "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (tdk)
FT      from B. subtilis"
FT      40543..41787
FT      cds      /*tag- l
FT      /label- MG035
FT      /note- "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hiss)
FT      from Mycobacterium leprae"
FT      complement (44751..46277)
FT      /*tag- m
FT      /label- MG038
FT      /note- "Previously identified as MORF-20105, the
FT      encoded protein shows 46.83 percentage
FT      identity to glycerol kinase (glpk)
FT      from E. coli"
FT      49377..49643
FT      cds      /*tag- n
FT      /label- MG039
FT      /note- "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (gdt2) from S. cerevisiae"
FT      49377..49643
FT      /*tag- o
FT      /label- MG041
FT      /note- "The encoded protein shows 48.86 percentage
FT      identity to phosphohistidinoprotein-hexose
FT      phosphotransferase (psh) from Mycoplasma
FT      capricolium"
FT      50060..51520
FT      cds      /*tag- p
FT      /label- MG042
FT      /note- "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      51525..52382
FT      cds      /*tag- q
FT      /label- MG043
FT      /note- "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (pots) from E. coli"
FT      52366..53220
FT      /*tag- r
FT      /label- MG044
FT      /note- "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein c (potc) from E. coli"
FT      54658..55605
FT      cds      /*tag- s
FT      /label- MG046
FT      /note- "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to sialoglycoprotease (gcp)
FT      from Pasteurella haemolytica"
FT      complement (56970..58310)
FT      /*tag- t
FT      /label- MG048
FT      /note- "Previously identified as MORF-19834,
FT      MORF-20114 and MORF-20115, the encoded protein
FT      shows 43.02 percentage identity to signal
FT      recognition particle protein (ifh) from B.
FT      subtilis"
FT      58117..59079
FT      /*tag- u
FT      /label- MG049
FT      /note- "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      59083..59754
FT      cds      /*tag- v
FT      /label- MG050
FT      /note- "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deoc) from Mycoplasma pneumoniae"
FT      complement (64898..65731)
FT      /*tag- w
FT      /label- MG056
FT      /note- "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185..99 from B. subtilis"
FT      81047..82597
FT      cds      /*tag- x
FT      /label- MG057
FT      /note- "Previously identified as MORF-20123, the
FT      encoded protein shows 38.90 percent
FT      identity to the protein disclosed in
FT      GB:D26185..104 from B. subtilis"
FT      81047..82597
FT      /*tag- y
FT      /label- MG067
FT      /note- "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (spase) from Staphylococcus aureus"
FT      91065..91919
FT      /*tag- z
FT      /label- MG070
```

FT /note= "Previously identified as MORF-20136, the
 FT encoded protein shows 34.8 percentage
 FT identity to ribosomal protein S2 (rps2)
 FT from *Spirulina plantensis*"
 FT 103104..104324
 FT cds /tag= aa
 FT /label= MG077
 FT /note= "Previously identified as MORF-20140, the
 FT encoded protein shows 28.05 percentage
 FT identity to oligopeptide transport system

... Note: remainder of annotations omitted.

Query Match 48.0%; Score 96; DB 27; Length 580073;
 Best Local Similarity 43.2%; Pred. No. 3.07e+01;
 Matches 19; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 Db 387537 tggttaccattgtctgcactcttctgtgtgcaaac 387580
 Cp 50 TGRITNGCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCC 7

RESULT 5
 ID 046579 standard; DNA; 359 BP.
 AC 046579;
 DT 22-DEC-1993 (first entry)
 DE Active site polynucleotide fragment.
 KM Active site; active region; prediction; secondary structure;
 OS energy value; engineering; ss.
 PN Synthetic.
 PD 05155899-A.
 PE 22-JUN-1993.
 PR 23-AUG-1991; 212284.
 PA (SAGA) SAGA1 CHEM RES CENTRE.
 DR WPI; 93-232353/29.
 PT Prediction of the active site in physiologically active
 PT polypeptide - by determ. of sec. structure energy values of
 PT partial regions along prim. sequence of physiologically active
 PT polynucleotide, etc.
 PS Disclosure: Page 23-24; 43pp; Japanese.
 CC This sequence represents a polynucleotide which was used to
 CC demonstrate the method of the invention. This oligomer represents
 CC an active site/region and the method of the invention may be used to
 CC predict the active site of a polynucleotide by determining the energy
 CC values of the secondary structure of various partial regions of the
 CC polynucleotide to be tested and the primary sequence as well. At
 CC least one partial region should show a local peak energy value
 CC significantly higher or lower than the average standard energy value.
 CC This method may be used in the elucidation of the mechanism of
 CC action of polynucleotides or polypeptides, and in the engineered
 CC improvement of such sequences.
 SQ Sequence 359 BP; 69 A; 87 C; 110 G; 93 T;
 Query Match 47.5%; Score 95; DB 7; Length 359;
 Best Local Similarity 44.4%; Pred. No. 3.67e+01;
 Matches 16; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
 Db 52 caggtgcgcgaagccacacgtactatgaacc 87
 Oy 14 CNGNGCNCMGNAARWSGNCMGNAARITNGCNAAYC 49
 RESULT 6
 ID 060810 standard; DNA; 297 BP.
 AC 060810;
 DT 16-MAR-1994 (first entry)
 DE Human brain Expressed Sequence Tag EST00914.
 KM Gene transcription product; genetic markers; tagging; in vivo;
 OS Homo sapiens.
 PN M09316178-A.
 PD 19-AUG-1993.

PF 12-FEB-1993; U01294.
 PR 12-FEB-1992; US-837195.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Adams MD, Moreno RF, Venter CJ.
 DR WPI; 93-272882/34.
 PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 PS Example 4; Page 394; 500pp; English.
 CC The Expressed Sequence Tag was isolated from a human brain cDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prep. of antisense sequences, probes and constructs.
 CC EST00914 has a "poor" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also 059041-Q61440.
 SQ Sequence 297 BP; 126 A; 44 C; 45 G; 82 T;

Query Match 45.5%; Score 91; DB 8; Length 297;
 Best Local Similarity 43.6%; Pred. No. 7.41e+01;
 Matches 17; Conservative 7; Mismatches 15; Indels 0; Gaps 0;
 Db 170 gattcactgtgtaagtaaatgtacacacaccttcaa 208
 Oy 8 GNTTYACNGCNCMGNAARWSGNCMGNAARITNGCNA 46

RESULT 7
 ID T90224 standard; DNA; 1302 BP.
 AC T90224;
 DT 08-MAY-1998 (first entry)
 DE Magnetospirillum sp. Maga DNA degenerate homologue.
 KM Maga; immobilisation; magnetic particle; magnetic bacteria; ss.
 OS Magnetospirillum sp.
 PN W09735964-A1.
 PD 02-OCT-1997.
 PE 27-MAR-1997; J01043.
 PR 16-MAY-1996; JP-146833.
 PR 27-MAR-1996; JP-097536.
 PA (DENK) TDK CORP.
 PI Kamiya S, Matsunaga T, Namba K;
 DR WPI; 97-489628/45.
 PT Proteins bound to magnetic particles - useful as immobilised enzymes
 PT and analytical reagents
 PS Disclosure: Page 47-48; 70pp; Japanese.
 CC This sequence encodes a novel degenerate homologue of the Maga protein,
 CC which is used in a method of producing magnetic particles for the
 CC immobilisation of various substances. These particles may be used as
 CC substrates for immobilising enzymes, DNA carriers, a protein biosynthesis
 CC system, analytical reagents or to assay antigens or other substances. The
 CC amount of a substance in a sample can be measured using the products. The
 CC immobilised enzymes are easily prepared by culturing transformed magnetic
 CC bacteria, and isolating the magnetic particles.
 SQ Sequence 1302 BP; 143 A; 118 C; 259 G; 216 T;
 Query Match 45.5%; Score 91; DB 38; Length 1302;
 Best Local Similarity 26.5%; Pred. No. 7.41e+01;
 Matches 13; Conservative 11; Mismatches 25; Indels 0; Gaps 0;
 Db 999 rathgngartlywsntlytngngaracngngaargcngtaar 1047
 Cp 49 GRTTNGCNARYTTNCKNGCNSWYTTNCKNGCNCNGTRANCCGCAA 1
 RESULT 8
 ID 020264 standard; CDNA; 1606 BP.
 AC 020264;
 DT 25-MAR-1992 (first entry)
 DE Mec-4 gene.
 KM Neurodegeneration; cell death; nematode; Huntington's; Lou Gehrig's;
 KM Alzheimer's; roundworm; ss.

OS	Caenorhabditis elegans.	Location/Qualifiers
FT	Key	14..1495
FT	cds	/*tag= a
FT	polya_site	1567..1606
FT		/*tag= b
PN	W0919007-A.	
PD	12-DEC-1991.	
PF	30-MAY-1991; U03826.	
PR	30-MAY-1990; US-530986.	
PA	(UYCO-) COLUMBIA UNIV NEW Y.	
P1	Chalfie M, Wolinsky E, Driscoll M;	
DR	WPI: 92-007488/01.	
DR	P-PSDB: R20107.	
PT	DNA sequences involved in neuronal degeneration - used for	
P1	detecting, diagnosing, treating and preventing degenerative	
PT	diseases.	
PS	Claim 10: Fig 14: 57pp: English.	
CC	The mec-4 (mechanosensory abnormal) gene encodes a protein associa-	
CC	ted with the late-onset degeneration of a specific set of six touch	
CC	receptor neurons. Strains of C. elegans having mutations in the	
CC	mec-4 gene (el611, u214, u231) have been identified in a screen for	
CC	touch sensitivity. No description of the cloning method is provid-	
CC	ed. The DNA may be useful for detecting, diagnosing, preventing and	
CC	treating degenerative diseases such as Alzheimer's disease, amy-	
CC	otrophic lateral sclerosis and Huntington's disease.	
CC	See also Q20263.	
SO	Sequence 1606 BP; 500 A; 312 C; 360 G; 434 T;	
Query Match	45.5%; Score 91; DB 3; Length 1606;	
Best Local Similarity	43.3%; Pred. No. 7.41e+01;	
Matches	13; Conservative 8; Mismatches 9; Indels 0; Gaps 0;	
Db	1308 gatttgcactctcagcagatttggtg 1337	
Cp	49 GRTTGCNARTTNTCKNCNSMYTTCKNG 20	
RESULT	9	
ID	Q38934 standard; cDNA: 1607 BP.	
AC	Q38934;	
DT	20-JUL-1993 (first entry)	
DE	mec-4 gene.	
KW	deg-1; mec-4; nerve cells; neuronal cells; neurodegenerative;	
OS	disease; drug screening; testing; animal model.	
KM	Caenorhabditis elegans.	
FT	Key	Location/Qualifiers
FT	cds	15..1498
FT		/*tag= a
PN	US5196333-A.	
PD	23-MAR-1993.	
PF	30-MAY-1990; 530968.	
PR	30-MAY-1990; US-530968.	
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
P1	Chalfie M, Driscoll M, Wolinsky E;	
DR	WPI: 93-116796/14.	
DR	P-PSDB: R33890.	
PT	Isolated nucleic acid molecule encoding wild type free living	
PT	nematode protein - comprises C. elegans deg-1 gene with specific	
PT	DNA sequence and mutated for neuronal cell degeneration	
PS	Claim 5: Fig 9: 35pp: English.	
CC	This sequence represents the mec-4 gene from C. elegans. Mutations	
CC	in the deg-1 (Q38933) and mec-4 genes cause mature functioning nerve	
CC	cells in C. elegans to die. This can provide an animal model to	
CC	study the causes of neurodegenerative diseases, and to cause the	
CC	degeneration of a diseased human cell. The C. elegans strains can	
CC	also be used to screen drugs to identify those which prevent or	
CC	decrease neuronal degeneration.	
SO	Sequence 1607 BP; 501 A; 312 C; 360 G; 434 T;	
Query Match	45.5%; Score 91; DB 6; Length 1607;	
Best Local Similarity	43.3%; Pred. No. 7.41e+01;	
Matches	13; Conservative 8; Mismatches 9; Indels 0; Gaps 0;	

```

Db 1308 gatttgaactgtctagccattctgtg 1337
      ||| | ||| | | | ||| ||| |
Cc 49 GRTTGCNARTTTCNKGCSNWTTCCKNG 20

RESULT 10
ID VI6515 standard; DNA; 3471 BP.
AC VI6515;
DE 11-JUN-1998 (first entry)
   DNA encoding a Bacillus thuringiensis toxin designated 86Bb1(a).
   Toxin; lepidopteran pest; control; Agrotis ipsilon; black cutworm;
   Heliothis virescens; Helicoverpa zea; ss.
OS Bacillus thuringiensis.
FH Key
FT CDS
   Location/Qualifiers
   ..1..3471
   /*tag= a

FN MO9800546-A2.
PD 08-JAN-1998.
PE 01-JUL-1997; U11658.
PR 01-JUL-1996; US-674002.
PA (MYCO ) MYCOGEN CORP.
PI Marva KE, Schmepe HE, Stockhoff BA, Walz M, Wicker C;
DR WPI: 98-086971/08.
P-PSDB: W46856.
PT New isolated Bacillus thuringiensis isolate(s) - used to obtain
   genes encoding toxins which are active against lepidopteran pests
   such as the Black cutworm
   Example 5; Pages 86-88; 183pp: English.
PS The present sequence is isolated from a Bacillus thuringiensis isolate
   CC (P866Bb1). It encodes a toxin designated 86Bb1(a) which is active
   CC against lepidopteran pests. The toxin isolates can be used for the
   CC control of lepidopteran pests such as Agrotis ipsilon (black cutworm),
   CC Heliothis virescens and Helicoverpa zea. PCR primers and probes can be
   CC derived from the polynucleotide encoding the toxin and used for the
   CC amplification and detection of other toxin-encoding sequences.
   Sequence 3471 BP; 1144 A; 591 C; 734 G; 1002 T;

Query Match 45.5%; Score 91; DB 40; Length 3471;
Best Local Similarity 43.3%; Pred. No. 7.41e+01;
Matches 13; Conservative 8; Mismatches 9; Indels 0; Gaps 0

Db 2304 gctagcaagtgcagagaattaccacac 2333
      ||| | ||| | | | ||| | |||
Cc 19 GCNMGNARMSNGCMGNARTTGCNCAAY 48

RESULT 11
ID VI68003 standard; DNA; 36335 BP.
AC VI68003;
DE 26-MAR-1996 (first entry)
   Ad2-/ORF6/PSK nucleotide sequence.
   Recombinant adenovirus; Ad2/CFR-1; adenovirus 2 serotype; Ela; E1b;
   viral replication; gene expression; gene therapy; cystic fibrosis;
   cysic fibrosis transmembrane conductance regulator; CFR;
   KW promoter; E3; p19; MHC; class 1; viral latency; pulmonary airway; ds.
OS Synthetic.
FH Key
FT misc-feature
   Location/Qualifiers
   ..12915..36335
   /*tag= a
   /note= "Represents residues 10676-34096 of Ad2-E4/ORF6"
   35069..35973
   /*tag= b
   /note= "Represents residues 33178-34082 of Ad2"
   12915..35054
   /*tag= c
   /note= "Represents residues 1-32815of Ad2"
   28478..28790
   /*tag= d
   /number= 1
   /note= "33K protein"
   28791..28992
   /*tag= e

```

```

FT      /number-1
FT      28993..29366
FT      /tag- f
FT      /number-2
FT      /note="33K protein"
FT      13279..14526
FT      /tag- g
FT      /product= 52,55K protein
FT      14547..16304
FT      /tag- h
FT      /product= iIra protein
FT      16331..16336
FT      /note="peripentonal hexon-associated protein"
FT      /tag- i
FT      /note="Major late mRNA L1 poly A signal (putative)"
FT      16390..18105
FT      /tag- j
FT      /product= Penton protein
FT      /note="Viron component IIT"
FT      18112..18708
FT      /tag- k
FT      /product= Pro-VII protein
FT      /note="Precursor to major core protein"
FT      18778..19887
FT      /tag- l
FT      /product= pV protein
FT      /note="minor core protein"
FT      20188..20193
FT      /tag- m
FT      /note="major late mRNA L2 poly-A signal (putative)"
FT      20240..20992
FT      /tag- n
FT      /product= pVI protein
FT      /note="Hexon associated precursor"
FT      21077..23983
FT      /tag- o
FT      /product= Hexon protein
FT      /note="Viron component IIT"
FT      24657..24662
FT      /tag- p
FT      /note="Major late mRNA L3 poly-A signal (putative)"
FT      complement (244729..26318)
FT      /tag- q
FT      /product= DBP protein
FT      /note="DNA binding or 72K protein"
FT      26347..26764
FT      /tag- r
FT      /product= 100K protein
FT      /note="Hexon assembled"
FT      29454..30137
FT      /tag- s
FT      /product= pVIII protein
FT      /note="Hexon-associated precursor"
FT      30444..30449
FT      /tag- t
FT      /note="major late mRNA L4 poly-A signal (putative)"
FT      31051..31530
FT      /tag- u
FT      /product= E3 19K protein
FT      /note="Glycosylated membrane protein"
FT      31707..32012
FT      /tag- v
FT      /product= E3 11.6K protein
FT      32008..32013
FT      /tag- w
FT      /note="E3-1 mRNA poly-A signal (putative)"
FT      33081..33086
FT      /tag- x
FT      /note="E3-2 mRNA poly-A signal (putative)"
FT      35013..35018
FT      /tag- y
FT      /note="major late mRNA L5 poly-A signal (putative)"
FT      1..12914
FT      misc-feature

```

```

FT      /tag- z
FT      /note="Represents residues 1-12914 of pad2/PGK-CFTR"
FT      380..914
FT      /tag- aa
FT      /note="pgk promoter"
FT      1011..5453
FT      /tag- ab
FT      /product= CFTR
FT      MO9412649-A.
FT      09-JUN-1994.
FT      02-DEC-1993; U11667.
FT      03-DEC-1992; US-985478.
FT      01-OCT-1993; US-130682.
FT      13-OCT-1993; US-136742.
FT      (GENZ) GENZYME CORP.
FT      PI      Armentano D, Couture LA, Gregory RJ, Smith AE;
FT      WPT: 94-20027/24.
FT      DR      WPT: 94-20027/24.
FT      PT      Adeno:Virus-based gene therapy vectors - esp. useful for gene
FT      PS      therapy of cystic fibrosis
FT      PS      Example 15, Page 84-95; 167pp; English.
FT      CC      This sequence represents the nucleotide sequence of the second generation
FT      CC      adenovirus vector named Ad2-ORF6/PGK-CFTR. This virus lacks E1 and in its
FT      CC      place contains a modified transcription unit with the phosphoglycerate
FT      CC      kinase (pgk) promoter and a poly A addition site flanking the cystic
FT      CC      fibrosis transmembrane conductance regulator (CFTR) cDNA. The PGK
FT      CC      promoter is only if moderate strength but it is long lasting and is not
FT      CC      subject to shut off. The E4 region of the vector has also been modified
FT      CC      in that the whole sequence has been removed and replace by ORF6, the
FT      CC      only E4 gene essential for growth of adenovirus (Ad) in tissue culture.
FT      CC      The DNA construct comprises a full length copy of the Ad2 genome from
FT      CC      which the early region 1 genes (E1 genes) have been deleted and replaced
FT      CC      by an expression cassette encoding CFTR. The expression cassette
FT      CC      includes the promoter for PGK and a poly-A addition signal from the
FT      CC      bovine growth hormone gene. The Ad2-ORF6/PGK-CFTR construct differs
FT      CC      from that given in Q68002 (Ad2/CFTR-1), in that the latter utilises the
FT      CC      endogenous E1a promoter, has no poly-A addition signal directly
FT      CC      downstream of CFTR and retains an intact E4 region. This adenovirus may
FT      CC      be administered to the pulmonary airways in the gene therapy of cystic
FT      CC      fibrosis.
FT      SQ      Sequence 36335 BP; 8597 A; 10000 C; 9786 G; 7952 T;
FT
FT      Query Match 45.5%; Score 91; DB 16; Length 36335;
FT      Best Local Similarity 42.5%; Pred. No. 7.41e+01;
FT      Matches 17; Conservative 8; Mismatches 15; Indels 0; Gaps 0;
FT
FT      Db 34861 acttacattacactgaatgacatgacacagaa 34900
FT      Cp 40 RYTNCKKNGCNSWTTTNCNGCNCNGTIRANCCNCCRAA 1
FT
FT      RESULT 12
FT      ID T10156 standard; DNA; 3872 BP.
FT      AC T10156;
FT      DT 30-MAR-1996 (first entry)
FT      DE Sockeye salmon growth hormone type 2 gene OnGH2.
FT      KW Transgenic fish; salmonid; sockeye salmon; coho salmon;
FT      KW Atlantic salmon; vector; growth hormone type 1; somatotropin;
FT      KW ss.
FT      OS Oncorhynchus nerka.
FT      FH Key Location/Qualifiers
FT      FT promoter 1..288
FT      FT /tag- a
FT      FT tata_signal 260..264
FT      FT /tag- b
FT      FT exon 289..363
FT      FT /tag- c
FT      FT /label= Exon-1
FT      FT /codon_start= 354..356
FT      FT intron 364..833
FT      FT /tag- d
FT      FT /label= Intron-A
FT      FT exon 834..973
FT      FT /tag- e

```


CC expression of the phosphorylated kinase insert domain of a growth
 CC factor receptor kinase eg. platelet-derived growth factor receptor.
 SQ Sequence 4281 BP; 1075 A; 1170 C; 1121 G; 915 T;

Query Match 45.0%; Score 90; DB 9; Length 4281;
 Best Local Similarity 40.5%; Pred. No. 8.82e+01;
 Matches 17; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

DB 14 tgaagggttcacagccacgaatgcacgtatgaacacagc 55
 :| | | | | | | | | | | | | | | | | | | | | |
 QY 3 YGNGGNTTYACNGNCNMGNARMSNGMGAARYTNC 44

RESULT 15
 ID T97221 standard; DNA; 24417 BP.
 AC T97221;
 DT 27-APR-1998 (first entry)
 DE Pseudomonas aeruginosa wbp gene cluster and flanking DNA.
 KW O antigen; HmD; Wzz; PsdA; WbPA; PsbA; PsbC; WbPC;
 KW PabD; WbPD; PsbE; WbPE; Rfc; Wzy; PsbF; HsH; HsF; PsbG;
 KW WbPG; WbPH; PsbH; WbPI; PsbJ; WbPJ; PsbK; WbPK; PsbL; WbPL;
 KW PabM; WbPM; PsbM; WbPN; WbPN; ultraviolet repair; histidine;
 KW infection; diagnosis; lipopolysaccharide; B-band gene cluster;
 KM pV100; ss.
 OS Pseudomonas aeruginosa PAO1.
 FH Location/Qualifiers
 FT Key
 FT CDS
 FT 3..479
 FT /tag- a
 FT /product- Rol (Wzz)
 FT 1286..2596
 FT /tag- b
 FT /product- PsbA (WbPA)
 FT 2670..3620
 FT /tag- c
 FT /product- PsbB (WbPB)
 FT 3277..5577
 FT /tag- d
 FT /product- PsbC (WbPC)
 FT 5574..6065
 FT /tag- e
 FT /product- PsbD (WbPD)
 FT 6151..6981
 FT /tag- f
 FT /product- PsbE (WbPE)
 FT 7235..8551
 FT /tag- g
 FT /product- Rfc (Wzy)
 FT 8059
 FT /tag- h
 FT /note- "base 8059 may be G"
 FT 8548..9498
 FT /tag- i
 FT /product- PsbF (WbPF)
 FT 10032..10388
 FT /tag- j
 FT /product- HsH
 FT 10388..11143
 FT /tag- k
 FT /product- HsF
 FT 11281..12411
 FT /tag- l
 FT /product- PsbG (WbPG)
 FT 12427..13548
 FT /tag- m
 FT /product- PsbH (WbPH)
 FT 13545..14633
 FT /tag- n
 FT /product- PsbI (WbPI)
 FT 14651..15892
 FT /tag- o
 FT /product- PsbJ (WbPJ)
 FT 15889..16851
 FT /tag- p

FT /product- PsbK (WbPK)
 FT 16911..17822
 FT /tag- q
 FT /product- PsbL (WbPL)
 FT 17935..19144
 FT /tag- r
 FT /note- "IS1209(PA)"
 FT 18032..19141
 FT /tag- s
 FT /note- "IS407"
 FT 19678..21675
 FT /tag- t
 FT /product- PsbM (WbPM)
 FT 22302..23693
 FT /tag- u
 FT /product- PsbN (WbPN)
 FT 23704..24417
 FT /tag- v
 FT /product- WbTB
 FT CDS
 FT MO9741234-A2.
 FT 06-NOV-1997.
 FT 30-APR-1997; CA0295.
 FT 27-FEB-1997; US-039473.
 FT 30-APR-1996; US-016510.
 FT (UYGU-) UNIV GUELPH.
 FT PI Burrows L, Charter D, De Kievit T, Lam JS;
 FT WPI; 97-549736/50.
 FT P-PSbB; W37349-67.
 FT Pseudomonas aeruginosa B-band lipopolysaccharide gene cluster -
 FT used for diagnosis of P. aeruginosa infection
 FT claim 2: Page 92-104; 195pp: English
 FT This novel isolated Pseudomonas aeruginosa B-band gene cluster
 FT includes the novel genes rol (wzz), psbA (wbPA), psbB (wbPB),
 FT psbC (wbPC), psbD (wbPD), psbE (wbPE), rfc (wzy), psbF (wbPF),
 FT psbG (wbPG), psbH (wbPH), psbI (wbPI), psbJ (wbPJ), psbK (wbPK),
 FT psbM (wbPM) and psbN (wbPN) involved in the synthesis and assembly
 FT of B-band lipopolysaccharide (i.e. O-antigen). wvB involved in
 FT ultraviolet repair, and hsh and hsf involved in histidine
 FT synthesis. It comprises the entire 24416 bp insert of cosmid clone
 FT pV100. Also claimed are: purified proteins (see W37349-66) encoded
 FT by these genes; transformed host cells including a recombinant
 FT molecule comprising nucleic acids from the B-band gene cluster; and
 FT monoclonal or polyclonal antibodies that specifically recognise the
 FT B-band lipopolysaccharide. The antibodies can be used to detect P.
 FT aeruginosa in a sample, while probes and primers can be used to
 FT detect the presence of a B-band gene cluster nucleic acid in a
 FT sample. Both of these claimed methods can be used to diagnose P.
 FT aeruginosa infections.
 FT Sequence 24417 BP; 4991 A; 5938 C; 7165 G; 6323 T;
 SQ

Query Match 45.0%; Score 90; DB 38; Length 24417;
 Best Local Similarity 44.1%; Pred. No. 8.82e+01;
 Matches 15; Conservative 7; Mismatches 12; Indels 0; Gaps 0;
 DB 1056 gccacactgctgcaggttcacgaacacgaatga 1089
 :| | | | | | | | | | | | | | | | | | | | | |
 CP 44 GCNARITTNCKNGCNSWTTTNCKNGCNGTRRA 11

Search completed: Tue Apr 27 10:17:34 1999
 Job time : 33 secs.

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd.

(TM)

MPsrch_tpm n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Tue Apr 27 10:14:22 1999: MasPar time 129.38 Seconds

Tabular output not generated.

Title: >US-09-011-797-2
Description: (1-17) from US09011797.pep
Perfect Score: 200
N.A. Sequence: 1 TTYGGNGNTTACNGGNGC.....CNMGNAARYNCNAAAYCAR 51
Comp: AACCCNCCNAATGNCNCNC.....GNKNTTYRANCGNTTGTGT

Scoring table: TABLE bkttranslated2
Gap 40

Nmatch STD: Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-est56
Database: 1:em-est1 2:em-gss1 3:em-gss2 4:em-gss3
genbank-est109
5:gb-est1 6:gb-est10 7:gb-est11 8:gb-est12 9:gb-est13
10:gb-est14 11:gb-est15 12:gb-est16 13:gb-est17
14:gb-est18 15:gb-est19 16:gb-est20 17:gb-est21
18:gb-est22 19:gb-est23 20:gb-est24 21:gb-est25 22:gb-est26
23:gb-est27 24:gb-est28 25:gb-est29 26:gb-est30 27:gb-est31
28:gb-est32 29:gb-est33 30:gb-est34

Statistics: Mean 44.884; Variance 55.405; scale 0.810

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	196	98.0	621	18	EST222659 Normalized r	1.14e-25
2	116	58.0	419	13	AK33111.s1 Soares test	1.74e-05
3	116	58.0	488	13	AA867614	1.74e-05
4	111	55.5	396	28	AA0196667	2.15e-04
5	111	55.5	410	15	AA971684	2.15e-04
6	111	55.5	520	15	AA062871	2.15e-04
7	111	55.5	767	14	AA950429	2.15e-04
8	106	53.0	290	9	W09332	2.15e-03
9	106	53.0	302	20	KMTSG63X	2.50e-03
10	106	53.0	358	26	B78873	2.50e-03
11	106	53.0	376	27	AA085207	2.50e-03
12	106	53.0	386	7	HUM073F02B Human fetal brain cDNA	2.50e-03

13	106	53.0	403	11	AA671628	V108h07.r1 Soares	mous	2.50e-03
14	106	53.0	428	7	AA462650	V108h09.r1 Soares	mous	2.50e-03
15	106	53.0	437	26	FR0002492	E_rubripes GSS sequenc		2.50e-03
16	106	53.0	452	15	AA1060712	ub42904.r1 Soares 2ndM		2.50e-03
17	106	53.0	482	24	AA273453	vc01f11.r1 Soares mous		2.50e-03
18	106	53.0	490	28	AA0129132	HS_3032_B1.G10 MR CIT		2.50e-03
19	106	53.0	496	7	AA542077	VK43f10.r1 Soares mous		2.50e-03
20	106	53.0	2004	28	AF004855	Mus musculus strain SM		2.50e-03
21	105	52.5	399	17	AA117839	uc41b05.r1 Soares mous		4.04e-03
22	105	52.5	455	8	AA236408	zr54g09.r1 Soares NHM		4.04e-03
23	101	50.5	274	15	AA1012447	EST206898 Normalized r		2.68e-02
24	101	50.5	346	28	AA0105662	HS_3050_A2_F08 MF CIT		2.68e-02
25	101	50.5	384	7	AA450986	V165h01.r1 Soares mous		2.68e-02
26	101	50.5	386	14	AA900662	U1-R-E0-d1-f-11-0-U1.s		2.68e-02
27	101	50.5	419	18	AA1178179	U1-R-E0-d1-f-11-0-U1.s		2.68e-02
28	101	50.5	421	6	C05751	Similar to N-terminal		2.68e-02
29	101	50.5	437	12	AA745642	HY69904.s1 NCI CGAP GC		2.68e-02
30	101	50.5	437	18	AA1136553	U1-R-C2p-nw-c-02-0-U1.		2.68e-02
31	101	50.5	448	15	AA932335	0060b05.s1 NCI CGAP Lu		2.68e-02
32	101	50.5	460	14	AA925364	U1-R-A1-ee-c-10-0-U1.s		2.68e-02
33	101	50.5	462	17	AA118108	u145c04.x1 Sugano mous		2.68e-02
34	101	50.5	472	10	AA424081	zve0b04.x1 Soares tota		2.68e-02
35	101	50.5	478	12	AA808517	oe55e03.s1 NCI CGAP Lu		2.68e-02
36	101	50.5	480	16	H44229	Yp04a08.r1 Homo sapien		2.68e-02
37	101	50.5	486	20	H98532	za59a05.r1 Homo sapien		2.68e-02
38	101	50.5	492	16	H28425	Y034a03.s1 Homo sapien		2.68e-02
39	101	50.5	495	14	AA878310	oe61f06.s1 NCI CGAP Lu		2.68e-02
40	101	50.5	522	23	AA104153	mp02c02.r1 Stratagene		2.68e-02
41	101	50.5	554	11	AA635307	nt57e09.s1 NCI CGAP Pr		2.68e-02
42	101	50.5	556	6	AA401353	zue8e03.s1 Soares test		2.68e-02
43	101	50.5	598	12	AA440425	LD15018.Sprine LD Dros		2.68e-02
44	101	50.5	605	18	AA1090348	oy18b06.s1 Soares_sene		2.68e-02
45	101	50.5	711	26	AG002612	Homo sapiens genomic D		2.68e-02

ALIGNMENTS

RESULT	1	LOCUS	621 bp	mrna	EST	08-OCT-1998
DEFINITION	EST222659	Normalized rat spleen.	Bento Soares	Rattus sp.	cDNA clone	
ACCESSION	AF178977	AF178977				
NID	g3729615					
KEYWORDS	EST.					
SOURCE	Rattus sp.					
ORGANISM	Rattus sp.					
REFERENCE	1 (bases 1 to 621)					
AUTHORS	Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,					
TITLE	Eukaryota: Metazoa: Chordata: Vertebrata: Mammalia: Eutheria;					
JOURNAL	Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat					
COMMENT	Other_ESTs: TC52405					
FEATURES	source					
FEATURES	1..621					
FEATURES	/organism="Rattus sp."					
FEATURES	/note="Organ: spleen; Vector: pTTT3Pac; Site_1: EcoRI;					
FEATURES	Site_2: NotI					
FEATURES	/db_xref="taxon:10118"					
FEATURES	/clone="RSPB94"					
FEATURES	/clone_1ib="Normalized rat spleen, Bento Soares"					
BASE COUNT	140 a	166 c	166 g	149 t		
ORIGIN						

ACCESSION		sequence.
NID	AA867614	
KEYWORDS	Q2963059	
SOURCE	EST.	
ORGANISM	house mouse.	
	Mus musculus	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
REFERENCE	Rodentia; Sciurognathi; Muridae; Mus.	
AUTHORS	1 (bases 1 to 488)	
	Marrar,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,	
	Giesel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,	
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,	
	Wetzel,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and	
	Waterston,R.	
TITLE	The WashU-HMI Mouse EST Project	
JOURNAL	Unpublished (1996)	
COMMENT	Contact: Marra M/Mouse Est Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. MGI:666762 Seq primer: -28ml3 rev2 EN from Amersham High quality sequence stop: 484. Location/Qualifiers 1..488 /organism="Mus musculus" /strain="C57BL/6J" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5' TGTTACCATCTGTGAGCGAGCGCCGGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo." /db_xref="taxon:10090" /clone="1264210" /clone_id="soares 2Nbxr" /sex="male" /tissue_type="Thymus" /dev_stage="4 weeks" /lab_host="DH10B"	
FEATURES		
SOURCE		
BASE COUNT	110 a 117 g 164 t 2 others	
ORIGIN		
Query Match	58.0%; Score 116; DB 13; Length 488;	
Best Local Similarity	48.5%; Pred. No. 1,74e-05;	
Matches	16; Conservative 8; Mismatches 8; Indels 0; Gaps 0;	
Dd	45 TTGGTTTCCTGCCTTCCTGCTGCTCTGCTCTGCTACG 77	
Cp	51 YTGRTTNGCNARYYTTCNKNCNSWYTTCKKNCG 19	
RESULT		
LOCUS	4 A0196667 396 bp DNA GSS 15-SEP-1998	
DEFINITION	CIT-HSP-2381B2.TF CIT-HSP Homo sapiens genomic clone 2381B2,	
ACCESSION	A0196667	
NID	93604029	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
	Primates; Catarrhini; Hominoidea; Homo.	

[illegible]

```

NCI (C6P7 GCB3) were mixed, and ss circles were used in vitro. Following HAP purification, this DNA was made as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/cclone="IMAGE:1583660"
/clone_1ld="Soares_NFL_T_GBC.S1"
/lab_host="DH10B"

BASE COUNT      113 a      88 c      86 g      123 t

ORIGIN
Query Match          55.5%; Score 111; DB 15; Length 410;
Best Local Similarity 50.0%; Pred.No. 2,15e-04;
Matches 19; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Db    144 GGCTTGACCGGGGCAAGAGTCTGTCTACAAGCTGC 181
      |||||.....:::..:..:..:..:..:..:..:..:..:
      7 GGNTTYACNGGCMGNARMSNCCMGNAMARYTNGC 44

RESULT 6
LOCUS       A1062871      520 bp      mRNA      EST      24-JUL-1998
DEFINITION GH02225.3prime GH Drosophila melanogaster head POT2 Drosophila
            melanogaster CDNA clone GH02225 3prime, mRNA sequence.
ACCESSION   A1062871
KEYWORDS    93338710
SOURCE      ESM.
ORGANISM    fruit fly.
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
            Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 520)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
JOURNAL     BDGP/HMI Drosophila EST Project
COMMENT     Unpublished (1997)

Contact: Harvey, D.
          G. M. Rubin-Molecular and Cell Biology
          University of California Berkeley
          539 LSA, Berkeley, CA 94720-3200, USA
          Fax: 510 643 9947
Email: http://fruitfly.berkeley.edu/ESM, estefruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone probably contains an inverted insert. The sequence has been trimmed and the T residues removed. hit genomic sequence DS02155
Plate: 22 row: C column: 1
High quality sequence stop: 408.
Location/Qualifiers
1..520
/organism="Drosophila melanogaster"
/note="Organ: head; Vector: POT2; Site:1; EcoRI; Site:2; XhoI: Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."
/db_xref="taxon:7227"
/clone="GH02225"
/clone_1lb="GH Drosophila melanogaster head POT2"
/dev="male and female"
/dev_stage="adult"
/lab_host="DH5 alpha"
/lab_host="DH5 alpha" 91 t

BASE COUNT      155 a      155 c      119 g      91 t

ORIGIN
Query Match          55.5%; Score 111; DB 15; Length 520;
Best Local Similarity 41.7%; Pred.No. 2,15e-04;
Matches 20; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

```

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	8	ATTCCAGACTTACGCGGATTTTGGGCGCAAAACCAA	55							
Cp	48	RTTNCNARITTTCKNCGNSWTTTCKNCGNCNCTRANCCNCCRAA	1							
RESULT	7	AA950429	767 bp	mRNA	EST	04-MAY-1998				
LOCUS		LD30207.5	prime LD Drosophila melanogaster	embryo	POT2	Drosophila				
DEFINITION		melanogaster cDNA clone LD30207	5prime, mRNA sequence.							
ACCESSION		AA950429								
NID		93112666								
KEYWORDS		EST.								
SOURCE		fruit fly.								
ORGANISM		Drosophila melanogaster								
REFERENCE		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
AUTHORS		1 (bases 1 to 767)								
TITLE		Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.								
JOURNAL		BDGP/HMI Drosophila EST Project								
COMMENT		unpublished (1997)								
FEATURES		Source								
FEATURES		Location/Qualifiers								
FEATURES		1..767								
FEATURES		/organism="Drosophila melanogaster"								
FEATURES		/note="Organ: embryo; Vector: POT2; Site: 1; EcoRI; Site: 2; XhoI; Sized fractionated cDNAs were directly ligated into								
FEATURES		POT2."								
FEATURES		/db_xref="taxon:7227"								
FEATURES		/clone="LD30207"								
FEATURES		/clone_11b="LD Drosophila melanogaster embryo POT2"								
FEATURES		/sex="male and female"								
FEATURES		/dev_stage="0 to 24 hours mixed stage embryonic"								
FEATURES		/lab_host="X11 Blue"								
BASE COUNT		225 a	222 c	189 g	131 t					
ORIGIN										
Query Match		55.5%	Score 111;	DB 14;	Length 767;					
Best Local Similarity		41.7%	Pred. No. 2.15e-04;							
Matches		20;	Conservative 10;	Mismatches 18;	Indels 0;	Gaps 0;				
Db	25	ATTCCAGACTTACGCGGATTTTGGGCGCAATCGCAAAACCAA	72							
Cp	48	RTTNCNARITTTCKNCGNSWTTTCKNCGNCNCTRANCCNCCRAA	1							
RESULT	8	W09332	290 bp	mRNA	EST	01-OCT-1997				
LOCUS		me7db03.r1	Soares mouse pMNP19.5	Mus musculus	cDNA clone 303821	5'				
DEFINITION		similar to gp:xl100	NADH-UBIQUINONE OXIDOREDUCTASE 75 KD SUBUNIT							
ACCESSION		W09332								
NID		91283649								
KEYWORDS		EST.								
SOURCE		house mouse.								
ORGANISM		Mus musculus								
REFERENCE		Eumaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
AUTHORS		1 (bases 1 to 290)								
AUTHORS		Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and								
AUTHORS		Waterson, R.								

```

TITLE                               The WashU-HHMI Mouse EST Project
JOURNAL                             Unpublished (1996)
COMMENT

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNLN ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 200581
Seq primer: mob.REGA+ET
High quality sequence stop: 280.

FEATURES
    source
        Location/Qualifiers
            1..290
                /organism="Mus musculus"
                /note="Vector: pT73D (Pharmacia) with a modified
                polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTACCAATCTGAGTGAGGGAGCGCCGCTTTTCTTTTTTTT 3'] ,
                double-stranded cDNA was size selected, ligated to Eco R
                adapters (Pharmacia), was digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT73 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by Bentinck
                Soares and M.Fatima Bonaldo. RNA was kindly provided by
                Dr. Minoru Ko (Wayne State University)."
                /db_xref="taxon:10090"
                /clone_id="303821"
                /clone_1bp="Soares mouse p3NM19.5"
                /dev_stage="19.5 dpc total fetus"
                /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      86 a          70 c          61 g          73 t
ORIGIN

Query Match           53.0%; Score 106; DB 9; Length 290;
Best Local Similarity 48.5%; Pred. NO.2.50e-03;
Matches 16; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

Db 14 GCAAGCAGTAGCCTTCGAAGCTGTAACGAG 46
||::|||||::||::|||||::||::|||||
Qy 19 GCNNGNAAARMSNGCMGNAARTYTGCAACAICAR 51

RESULT 9
LOCUS       MMTSG83X             302 bp            RNA            EST
DEFINITION Mouse expressed sequence tag (EST) mRNA Tsgr83x, mRNA sequence.
ACCESSION   X61885
NID         955024
KEYWORDS    EST; expressed sequence tag.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Murinae; Mus.
1 (bases 1 to 302)
Hoog/C.
Direct Submission
Submitted (30-AUG-1991) C. Hoog, Dept of MOL Genetics, Karolinska
Institute, Box 60400, 10401 Stockholm, SWEDEN
2 (bases 1 to 302)
Isolation of a large number of novel mammalian genes by a
differential cDNA library screening strategy
Nucleic Acids Res. 19 (22), 6123-6127 (1991)
9206457
JOURNAL     JOURNAL MEDLINE
MEDLINE
FEATURES
    source
        Location/Qualifiers
            1..302
                /organism="Mus musculus"
                /strain="BAIB/c"
                /db_xref="taxon:10090"

```

BASE COUNT	77 a	66 c	72 g	87 t
ORIGIN	/clone.lib="Lambda uni-zap II"			
Query Match	53.0%	Score 106;	DB 20;	Length 302;
Best Local Similarity	46.3%	Pred. No. 2,50e-03;		
Matches	19;	Conservative	8;	Mismatches 14;
			Indels	0;
			Gaps	0;
Db	127	TGCGTACCTAGTTTCGTGATCCATTTTCCTGCCCTCTGCA	167	
CP	50	TGRTTNCNARYTTTNCNKCNCNSWYTTNCKNCNGTAA	10	
RESULT	10			
LOCUS	B78873	358 bp	DNA	GSS
DEFINITION	C17978SF-81H4.TV C17978SK Homo sapiens genomic clone 81H4, genomic survey sequence.			
ACCESSION	B78873			
NID	92865896			
KEYWORDS	GSS.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
	Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE	Kim,U.-J., Adams,M.D. and Simon,M.I.			
AUTHORS	1 (bases 1 to 358)			
TITLE	Determination of clone end sequences of human Bacterial Artificial			
JOURNAL	Chromosomes			
COMMENT	Unpublished (1997)			
	Other-GSS: C17978SK-81H4.TP			
	Contact: Ung-Jin Kim			
	Caltech Genome Research Lab			
	California Institute of Technology			
	Division of Biology, MS 147-75, Pasadena, CA 91125, USA			
	Tel: 626 796 7066			
	Fax: 626 395 4901			
	Email: ungjash.tree.caltech.edu			
	Clones are available from Research Genetics (info@resgen.com). BAC			
	end search page:			
	http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html			
	Seq primer: T7			
	Class: BAC ends.			
FEATURES	Location/Qualifiers			
source	1..358			
	/organism="Homo sapiens"			
	/note="Vector: PBAC108L; Site_1: HindIII; Site_2: HindIII"			
	/db_xref="taxon:9606"			
	/clone="81H4"			
	/clone.lib="C17978SK"			
	/sex="Female"			
	/cell.type="Fibroblast"			
BASE COUNT	105 a	90 c	60 g	103 t
ORIGIN				
Query Match	53.0%	Score 106;	DB 26;	Length 358;
Best Local Similarity	50.0%	Pred. No. 2,50e-03;		
Matches	16;	Conservative	7;	Mismatches 9;
			Indels	0;
			Gaps	0;
Db	200	GGGCTCTTTCCGCTCCGCAATCAACCA	231	
CP	34	KNCGNSWYTTNCKNKCNCNGTAAACNCRCR	3	
RESULT	11			
LOCUS	A0085207	376 bp	DNA	GSS
DEFINITION	HS-2270.AL.E08.MR.C17 Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2270 Col-15 Row-1, genomic survey sequence.			
ACCESSION	A0085207			
NID	93454424			
KEYWORDS	GSS.			
SOURCE	human.			

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 376)
AUTHORS	Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Tricofco,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E. Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors Unpublished (1997)
JOURNAL	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 2270 Row: I Column: 15 Class: BAC ends High quality sequence stop: 376. Location/Qualifiers 1..376
FEATURES	/organism="Homo sapiens" /note="Organ: Sperm; Vector: pBelBAC1; BAC Clones In E-Coli DH10B" /db_xref="taxon:9606" /clone="Plate=2270 Col=15 Row=I" /clone_lib="CIT Approved Human Genomic Sperm Library D" /sex="male"
SOURCE	126 a 100 c 75 g 75 t
BASE COUNT	
ORIGIN	
Query Match	53.0%; Score 106; DB 27; Length 376;
Best local Similarity	39.2%; Pred No. 2,50e+03;
Matches	20; Conservative 11; Mismatches 20; Indels 0; Gaps 0;
Db	171 TTGGGCTGCCTCAAGTGTCAAGATCGCTACACGTCGCCACTAG 221 : : : : : QY 1 TTGGNGNTTYACNGCGMGNARNSNCGNMARYTGCAAAACAR 51
RESULT	12
LOCUS	HUM073F02B 386 bp mRNA EST 21-MAY-1996
DEFINITION	Human fetal brain cDNA 5'-end GEN-073F02, mRNA sequence.
ACCESSION	D52294
NID	9852350
KEYWORDS	EST; EST(expressed sequence tag); Human fetal brain; similar to known (May 29, 1995).
SOURCE	Homo sapiens CDNA to mRNA, clone_11d:Ciontech human fetal brain polyA+ mRNA (#6535).
ORGANISM	Homo sapiens Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 386)
AUTHORS	Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shinada,Y., Shinozuka,H., Takachihi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H., Shin,S. and Nakamura,Y.
TITLE	Unpublished
JOURNAL	Unpublished (1995)
REFERENCE	2 (bases 1 to 386)
AUTHORS	Fujiwara,T.
TITLE	Direct Submission
JOURNAL	Submitted (30-MAY-1995) to the DDBJ/EMBL/GenBank databases. Tsutomu Fujiwara, Otsuka GEN Research Institute,Otsuka Pharmaceutical Co.,Ltd; 463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima 771-01, Japan (Tel:0886-65-2888, Fax:0886-37-1035) Submitted (30-May-1995) to DDBJ by: Tsutomu Fujiwara Otsuka GEN Research Institute
COMMENT	

BASE COUNT 139 a 83 c 84 g 122 t
 ORIGIN

Query Match 53.0%; Score 106; DB 7; Length 428;
 Best Local Similarity 48.5%; Pred. No. 2.50e-03;
 Matches 16; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

DB 35 GCAAGCACTGACCTGCAAGCTAGTAACACG 67
 19 GCNMGNAARMSGCMNGMNAARYTNCNAAACAR 51

RESULT 15
 LOCUS FR0002492 437 bp DNA GSS 27-FEB-1997
 DEFINITION F.rubripes GSS sequence, clone 004C10BC10, genomic survey sequence.
 ACCESSION Z86275
 MID 91883187
 KEYWORDS GSS; genome survey sequence.
 SOURCE Fugu rubripes.
 ORGANISM Fugu rubripes.
 Eukaryote; Mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae;
 Tetraodontidae; Fugu.
 1 (bases 1 to 437)
 Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranta,Y.,
 Williams,G. and Brenner,S.
 Direct Submission
 Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.mrc.ac.uk
 Vector: pBluescript II KS
 V_type: phagemid
 PRIMER: KS
 DESCR:
 One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source location/Qualifiers:
 1..437
 /organism="Fugu rubripes"
 /db_xref="taxon:31033"
 /clone_1lb="cosmid.004C10"
 /clone="004C10BC10"

BASE COUNT 114 a 114 c 94 g 108 t 7 others
 ORIGIN

Query Match 53.0%; Score 106; DB 26; Length 437;
 Best Local Similarity 51.1%; Pred. No. 2.50e-03;
 Matches 24; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

DB 17 CTGGTTGCACCTGCTGCTTTTACCGCACCAGTGAGCCAC 63
 51 YTGRTTNCNARYTNCNKGNCNWTTCNKGNCNCGTFAANCNC 5

Search completed: Tue Apr 27 10:16:42 1999
 Job time : 140 secs.

THIS PAGE BLANK (USPTO)

MIRAGE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Apr 16 14:21:22 1999; MasPar time 2.91 Seconds
94,600 Million cell updates/sec

Tabular output not generated.

Title: >US-09-011-797-3

Description: (1-17) from US09011797.pep

Perfect Score: 131

Sequence: 1 FSEFMROYLVLSMOSSO 17

Scoring table: PAM 150
Gap 15

Searched: 131922 segs, 16180660 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

a- geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 20.090; Variance 71.751; scale 0.280

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	100.0	17 24	W25163	Human oploid receptor	4.74e-06
2	60	45.8	700 23	W19992	Human CAMP used to id	6.50e+01
3	59	45.0	321 20	W04560	Carnation ACC oxidase	8.00e+01
4	59	45.0	433 21	W14001	Enolase protein.	8.00e+01
5	58	44.3	39 28	W44735	Residues 307-355 of c	9.83e+01
6	58	44.3	270 28	W44734	Residues 307-355 of c	9.83e+01
7	58	44.3	319 23	W19357	Ethylene synthase fto	9.83e+01
8	57	43.5	618 25	W26541	Trypanosoma cruzi ant	1.21e+02
9	57	43.5	3457 12	R62504	Large polypeptide seq	1.21e+02
10	56	42.7	273 22	W24665	H. pylori outer membr	1.48e+02
11	56	42.7	273 22	W20673	H. pylori outer membr	1.48e+02
12	56	42.7	273 22	W20673	H. pylori outer membr	1.48e+02
13	56	42.7	343 22	W20333	H. pylori cytoplasmic	1.48e+02
14	56	42.7	343 22	W24634	H. pylori cytoplasmic	1.48e+02
15	55	42.0	16 17	R77675	Glucosylase N-termin	1.81e+02
16	55	42.0	318 22	W09880	ACC oxidase GEF-1.	1.81e+02
17	55	42.0	345 1	R02218	Sequence encoding str	1.81e+02
18	55	42.0	466 19	R95688	Human GAP-SH3 domain	1.81e+02

RESULT ID	1	W25163	standard; peptide; 17 AA.
AC	W25163;		
DT	08-DEC-1997	(first entry)	
DE	Human oploid receptor-like 1 receptor ligand, nociceptin, fragment.		
KW	Oploid receptor-like 1; ORL1; ligand; promiscuity; stress;		
KW	hyperalgesia; locomotor activity; anxiety; neuroendocrine;		
KW	memory; attention; sensory perception; learning; homeostasis;		
KW	hypolgesia; nociceptin; endorphin; dynorphin A.		
OS	Synthetic.		
PN	W09707208-21.		
PD	27-FEB-1997.		
PF	14-AUG-1996; BE0087.		
PR	15-AUG-1995; US-002368.		
PA	(ULBR) UNIV LIBRE BRUXELLES.		
PI	Meunier J, Mollereau C, Parmentier M, Vassart G;		
DR	WPI: 97-165292/15.		
N-PSDB:	T79888.		
PT	Novel ligand for the oploid receptor-like receptor, nociceptin		
PT	has pro-nociceptive properties, useful for treating or preventing		
PT	diseases related to e.g. stress, hyperalgesia, locomotor activity,		
PT	etc.		
PS	claim 7, page 27, 48pp; English.		
CC	W25163-W25164 are fragments of a ligand to human oploid receptor-like 1		
CC	(ORL1) receptor, designated nociceptin, which resembles the endorphin		
CC	dynorphin A. Inhibitors of nociceptin or its DNA, e.g. antisense		
CC	sequences and antibodies can be used as new types of drugs in the		
CC	control of various behaviours or functions. The inhibitors can be used		
CC	to prevent or treat a disease related to hyperalgesia, neuroendocrine		
CC	secretion, stress, locomotor activity, anxiety, instinctive behaviour,		
CC	learning and memory, homeostasis, hyperalgesia, hypalgesia and/or		
CC	sensory perception.		
CC	Sequence 17 AA;		
SO	Sequence 17 AA;		

ALIGNMENTS

Query Match	100.0%	Score 131.	DB 24:	Length 17:
Best Local Similarity 100.0%;				
Matches 17; Conservative				
0; Mismatches 0;				
Indels 0;				
Gaps 0;				

Db 1 fsefmroylvlsmsq 17
|||||
OY 1 fsefmroylvlsmsq 17

Query Match	45.88;	Score 60;	DB 23;	Length 700;
Best Local Similarity	62.58;	Pred. No. 6.50e+01;		
Matches	5;	Conservative	3;	Mismatches 0;
				Indels 0;
				Gaps 0;
Db	328 fsdflrhy 335			
	: :			
OY	1 FSEFMR0Y 8			

Seq	Sequence	321 AA:
Query Match		Score 59; DB 20; Length 321;
Best Local Similarity	45.0%;	Pred. No. 8.0e-01;
Matches	6; Conservative	3; Mismatches 8; Indels 0; Gaps 0;
Db	287 fedymnlylkikfgeke 303	
Oy	1 FSEFMROYLVMSOSSO 17	

	Query Match	45.0 %;	Score 59;	DB 21;	Length 433;
	Best Local Similarity	31.3 %;	Pred. No.	8.00e+01;	
	Matches	5;	Conservative	8;	Mismatches 3; Indels 0; Gaps 0.
Dd	279 ygdgfvrdypvvsiedp	294			
OY	: : : : : : :				
	1 fsefmrkvlylsmcgs	16			

RESULT	5
ID	W44735 standard; peptide; 39 AA.
AC	W44735.
DT	18-MAY-1998 (first entry)
DE	Residues 317-355 of calpain I.
KW	Human; calpain I; antibody; detection; diagnosis; Alzheimer's disease.
OS	Synthetic.
OS	Homo sapiens.
PN	J08113596-A.
PD	07-MAY-1996.
PF	16-AUG-1995; 208740.
PR	17-AUG-1994; US-292080.
PA	(MCLE-) MCLEAN HOSPITAL CORP.
DR	WPI; 98-021953/03.
PT	Calpain-combining antibody - used for the diagnosis of Alzheimer's disease
PS	Claim 13; Page -, 11pp; Japanese.
CC	This sequence represents amino acids 317-355 of the human calpain I.
CC	The invention relates to an antibody which binds to amino acids 346-361
CC	(W44733), corresponding to the domain IIR of the protein, and can be
CC	used in the detection and diagnosis of Alzheimer's disease.
CC	(Note: this sequence is not specifically given in the specification but

CC 1s generated from a longer amino acid sequence (see W44734))
 SO Sequence 39 AA;

Query Match 44.3%; Score 58; DB 28; Length 39;
 Best Local Similarity 38.5%; Pred. No. 9.83e+01;
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 11 fdfmrefrltel 23
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSM 13

RESULT 6

ID W44734 standard; peptide: 270 AA.
 AC W44734;
 DT 18-MAY-1998 (first entry)
 DE Residues 307-576 of calpain I.
 KM Human; calpain I; antibody; detection; diagnosis; Alzheimer's disease.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT MISC-difference 144
 FT MISC-difference 261 /label= unknown
 FT MISC-difference 261 /label= unknown
 FT J0813596-A.
 DT 07-MAY-1996.
 PF 16-AUG-1995; 208740.
 PR 17-AUG-1994; US-292080.
 PA (MCLE-) MCLEAN HOSPITAL CORP.
 DR WPI: 98-021953/03.
 PT Calpain-combining antibody - used for the diagnosis of Alzheimer's
 PI disease
 PS Disclosure: Fig 11. 11pp; Japanese.
 CC This sequence represents amino acids 307-576 of the human calpain I.
 CC The invention relates to an antibody which binds to amino acids 346-361
 CC (W44733), corresponding to the domain III of the protein, and can be
 CC used in the detection and diagnosis of Alzheimer's disease.
 SO Sequence 270 AA;

Query Match 44.3%; Score 58; DB 28; Length 270;
 Best Local Similarity 38.5%; Pred. No. 9.83e+01;
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 21 fdfmrefrltel 33
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSM 13

RESULT 7
 ID W19537 standard; Protein: 319 AA.
 AC W19537;
 DT 16-SEP-1997 (first entry)
 DE Ethylene synthase from poplar plants induced by ozone.
 KM Ozone tolerant; atmospheric pollution; photochemical oxidant;
 OS Populus nigra.
 PN J09098785-A.
 PD 15-APR-1997.
 PF 06-OCT-1995; 284363.
 PR 06-OCT-1995; JP-284363.
 PA (TOYT) TOYOTA JIDOSHA KK.
 DR WPI: 97-275444/25.
 DR N-PSDB: T74024.
 PT Ozone inducible ethylene synthase gene derived from poplar plants -
 PT used for the production of ozone-tolerant plants
 PT Claim 7: Page 6-7; 9pp; Japanese.
 CC The present sequence represents a plant ethylene synthase, derived
 CC from poplar plants, which is induced by ozone. The gene encoding
 CC the ethylene synthase is used to develop a plant highly resistant
 CC to ozone as a major photochemical oxidant in atmospheric pollution.
 CC The plants use ozone for gene expression and in doing so reduce ozone
 CC levels in lower atmosphere levels.
 SO Sequence 319 AA;

Query Match 44.3%; Score 58; DB 23; Length 319;
 Best Local Similarity 29.4%; Pred. No. 9.83e+01;
 Matches 5; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 282 fddymklyaglkfagke 298
 | : | | | : | : | :
 QY 1 FSEFMROYLVLSMOSQ 17

RESULT 8

ID W26541 standard; Protein: 618 AA.
 AC W26541;
 DT 12-JAN-1998 (first entry)
 DE Trypanosoma cruzi antigen.
 KM Antigen; epitope; vaccine; protective immunity; Chagas disease;
 KM diagnosis; therapy; immunoassay.
 OS Trypanosoma cruzi Tulhman strain C2.
 PN M09718475-A1.
 PD 22-MAY-1997.
 PF 14-NOV-1996; U18624.
 PR 14-NOV-1995; US-557309.
 PA (CORI-) CORIXA CORP.
 PI Houghton RL, Lodes MJ, Reed SG, Skelky YAM.
 DR WPI: 97-289413/26.
 DR N-PSDB: T69167.
 PT Diagnosing Trypanosoma cruzi infection by detecting antibodies to
 PT novel antigens - which are useful in vaccines to provide protective
 PT immunity against Chagas' disease
 PS Disclosure: Page 88-91; 110pp; English.
 CC This polypeptide sequence comprises a full-length antigen of
 CC Trypanosoma cruzi, identified by sequencing a DNA clone (see T69167)
 CC obtained by screening a Trypanosoma cruzi genomic expression library
 CC with pools of sera from infected individuals. T. cruzi antigens
 CC (see W26530-41), or epitope-containing repeat sequences (see W19094-
 CC 102, W19079-86 and W26542-44) of native antigens, can be used in a
 CC variety of immunoassays for detecting T. cruzi infection in a
 CC blood, serum, plasma, saliva, cerebrospinal fluid or urine sample.
 CC The polypeptides are also useful in vaccines and pharmaceutical
 CC compositions for inducing protective immunity against Chagas
 CC disease. They can be produced by expression in transformed or
 CC transfectected host cells.
 SO Sequence 618 AA;

Query Match 43.5%; Score 57; DB 25; Length 618;
 Best Local Similarity 60.0%; Pred. No. 1.21e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 189 ylvymqgave 198
 | | | | | : | : | :
 QY 8 YLVLSMOSQ 17

RESULT 9
 ID R62504 standard; Protein: 3457 AA.
 AC R62504;
 DT 18-MAY-1995 (first entry)
 DE Large polyprotein sequence of Maize Chlorotic Dwarf Virus.
 KM MCDV; maize chlorotic dwarf virus; coat protein; plasmid vector;
 KM plant resistance; viral infection resistance; maize plant;
 KM maize dwarf mosaic virus; MDMV.
 OS Maize chlorotic dwarf virus.
 FH Key Location/Qualifiers
 FT region 439..829
 FT /note= "antigenic region of cp2"
 FT cleavage_site 896..897
 FT /label= dipeptide-cleavage_site
 FT /note= "may be used by animal picornavirus 3C
 FT proteases"
 FT peptide 897..1098
 FT /note= "putative cp3 region, one of the structural
 FT proteins"
 FT peptide 897..911

FT /note="N-terminus of cp3"
 FT 1099..1113
 FT /note="N-terminus of cp1"
 FT 1205..1483
 FT /note="antigenic region of cp1"
 FT 1098..1099
 FT /label= dipeptide, cleavage_site
 FT /note="may be used by animal picornavirus 3C
 FT proteases"
 PN W09421966-A.
 PD 29-SEP-1994.
 PF 22-MAR-1994; U03028.
 PR 24-MAR-1993; US-038768.
 PA (PION-) PIONEER HI-BRED INT INC.
 PA (USDA) US SEC OF AGRIC.
 PI McMullen MD, Roth BA, Townsend R;
 DR WPI: 94-317016/39.
 DR N-PSDB: Q74694.
 PT DNA encoding maize chlorotic dwarf virus proteins - used to
 PT provide plants with resistance to the virus and related viral
 PT infections
 PS Disclosure: Page 18-32; 40pp; English.
 CC This sequence shows the large polyprotein of the Maize Chlorotic
 CC Dwarf Virus (MCDV). When cleaved it contains 3 structural proteins,
 CC cp1, cp2 and cp3. Depending on the exact location of cp2, the MCDV
 CC genome, can encode upto 78 kD of protein 5' of the capsid proteins
 CC (for which there are no corresponding animal picornavirus protein).
 CC The DNA is used for imparting resistance to MCDV or viruses to which
 CC MCDV infection or resistance provides cross-resistance, including
 CC maize dwarf mosaic virus strain A. Any or all of the three coat
 CC protein genes from MCDV can be used to provide protection for plants.
 SQ Sequence 3457 AA;

Query Match 43.5%; Score 57; DB 12; Length 3457;
 Best Local Similarity 40.0%; Pred. No. 1.21e+02;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 1945 fmeflkmyaalymdn 1959
 Oy 1 FSEFMROYLVLSMOSS 15
 ID W24665 standard; Protein; 273 AA.
 AC W24665;
 DT 12-AUG-1997 (first entry)
 DE H. pylori outer membrane protein, 36126938.aa.
 KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
 KW detection; antisense; inhibition.
 OS Helicobacter pylori.
 PN W09719098-A1.
 PD 29-MAY-1997.
 PF 15-NOV-1996; U18542.
 PR 17-NOV-1995; US-561469.
 PA (ASTR) ASTRA AB.
 PI Smith DH;
 DR WPI: 97-298052/27.
 DR N-PSDB: T77483.
 PT Helicobacter pylori nucleic acid sequences and related proteins -
 PT used for diagnostics and therapeutics
 PS Claim 18; Page 179-180; 235pp; English.
 CC This sequence represents an H. pylori outer membrane protein.
 CC Helicobacter pylori has been strongly linked to chronic gastritis and
 CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC are used to evaluate compounds, especially activators or inhibitors of
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 CC sequence. The nucleic acid sequences, and corresponding proteins, are
 CC also useful for generating vaccines for immunising subjects against H.
 CC pylori or for use in detecting the presence of Helicobacter species in
 CC a sample. Antisense nucleic acid sequences of these sequences are
 CC used to inhibit expression of a gene from Helicobacter species. H.

CC pylori whole genomic DNA was isolated and nebulised to a median size of
 CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 CC complementary to the BstXI-cut PMP vectors, while the overhang is not
 CC self-complementary. Therefore the linkers will not concatemise nor
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
 CC were ligated to each of the 20 PMP vectors to construct a series of
 CC shotgun subclone libraries. The purified DNA samples were then
 CC sequenced.
 CC Note: The ORF/protein reference number for this sequence was obtained
 CC from the related specification, W09640893.
 SQ Sequence 273 AA;

Query Match 42.7%; Score 56; DB 22; Length 273;
 Best Local Similarity 35.7%; Pred. No. 1.48e+02;
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 2 efmkkfvalg11sa 15
 Oy 3 EFMROYLVLSMOSS 16
 ID W20465 standard; protein; 273 AA.
 AC W20465;
 DT 29-JUL-1997 (first entry)
 DE H. pylori outer membrane protein, 36126938.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglinch OT, Smith D, Mellgaard BJ;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67803.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 635; 1481pp; English.
 CC The present sequence is a H. pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 53679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 273 AA;

Query Match 42.7%; Score 56; DB 22; Length 273;
 Best Local Similarity 35.7%; Pred. No. 1.48e+02;
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 2 efmkkfvalg11sa 15
 Oy 3 EFMROYLVLSMOSS 16
 ID W20673 standard; protein; 277 AA.
 AC W20673;
 DT 22-JUL-1997 (first entry)

DE H. pylori outer membrane protein, 04c11617orf2.
 KM Cytoplasmic: vaccine: prevention; treatment: infection; identification;
 KM binding compound; bacterium: life cycle; activator; bacteria; inhibitor;
 KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996: U09122.
 PR 07-JUN-1995: US-487032.
 PR 01-APR-1996: US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR N-PSDB: T67926.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56: Page 1094: 1481pp: English.
 CC The present sequence is a H. pylori outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC Useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 277 AA;

Query Match 42.7%; Score 56; DB 22; Length 277;
 Best Local Similarity 35.7%; Pred. No. 1.48e+02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 6 efmkrtvalg11sa 19

QY 3 EFMROYLVLSMOSS 16

RESULT 13
 ID W20333 standard; protein: 343 AA.

AC W20333;

DT 29-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, 24816915.aa.

KM Vaccine: prevention; treatment: infection; identification;
 KM binding compound; bacterium: life cycle; activator; bacteria;
 KM inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KM cytoplasmic.

OS Helicobacter pylori.

PN W09640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996: U09122.

PR 07-JUN-1995: US-487032.

PR 01-APR-1996: US-630405.

PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;

DR WPI: 97-052306/05.

DR N-PSDB: T67772.

PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

PS Claim 61: Pages 524-525: 1481pp: English.

CC The present sequence is a Helicobacter pylori cytoplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically
 CC shearing the bacterial DNA. The sequences were analysed for ORF of

CC at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from
 CC various ORF were analysed for significant homology to other known
 CC or exported membrane proteins. Having identified and determined
 CC the sequences of interest, particular regions can be isolated from
 CC H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 343 AA;

Query Match 42.7%; Score 56; DB 22; Length 343;
 Best Local Similarity 63.6%; Pred. No. 1.48e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 fmeflkqrlv1 81

QY 1 FSEFMROYLV1 11

RESULT 14
 ID W24634 standard; protein: 343 AA.

AC W24634;

DT 11-AUG-1997 (first entry)

DE H. pylori cytoplasmic protein, 24816915.aa.

KM Chronic gastritis; duodenal ulcer disease; activator;
 KM inhibitor; bacterial life cycle; vaccine; immunisation; detection;
 KM antisense; inhibition; cytoplasmic.

OS Helicobacter pylori.

PN W09719098-A1.

PD 29-MAY-1997.

PF 15-NOV-1996: U18542.

PR 17-NOV-1995: US-561469.

PA (ASTR) ASTRA AB.

PI Smith DH;

DR WPI: 97-298052/27.

DR N-PSDB: T77452.

PT Helicobacter pylori nucleic acid sequences and related proteins -
 PT used for diagnostics and therapeutics

PS Claim 18; Pages 160-161: 235pp: English.

CC The present sequence is a Helicobacter pylori cytoplasmic
 CC protein.

CC H. pylori has been strongly linked to chronic gastritis and
 CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC are used to evaluate compounds, especially activators or inhibitors
 CC of bacterial life cycle, for the ability to bind an H. pylori
 CC nucleic acid sequence. The nucleic acid sequences, and
 CC corresponding proteins, are also useful for generating vaccines for
 CC immunising subjects against H. pylori or for use in detecting the
 CC presence of Helicobacter species in a sample. Antisense nucleic
 CC acid sequences of these sequences are used to inhibit expression of
 CC a gene from Helicobacter species. H. pylori whole genomic DNA was
 CC isolated and nebulised to a median size of 2000 bp. Purified DNA
 CC fragments were blunt-ended and ligated to unique BstXI-linker
 CC adapters in 100-1000 fold molar excess. These linkers are
 CC complementary to the BstXI-cut PMX vectors, while the overhang is
 CC not self-complementary. Therefore the linkers will not
 CC concatamerise nor will the cut vector re-ligate itself easily. The
 CC linker-adaptor inserts were ligated to each of the 20 PMX vectors
 CC to construct a series of shotgun subclone libraries. The purified
 CC DNA samples were then sequenced.

CC Note: The ORF/protein reference number for this sequence was
 CC obtained from the related specification, W09640893.

CC Sequence 343 AA;

SQ

Query Match 42.7%; Score 56; DB 22; Length 343;
 Best Local Similarity 63.6%; Pred. No. 1.48e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 71 fmeflkqrlv1 81

QY 1 FSEFMROYLV1 11

```

RESULT 15
ID R77675 standard; Peptide: 16 AA.
AC R77675;
DT 29-JUN-1996 (first entry)
DE Glucoamylase N-terminal secretion sequence.
KW glucoamylase; Yeast; Saccharomyces cerevisiae; production; starch;
   glucose; degradation; carbon source.
OS Arxula adeninivorans.
PN DE4425058-A.
PD 18-JAN-1996.
PF 15-JUL-1994; 425058.
PR 15-JUL-1994; DE-425058.
PA (PFLA-) INST PFLANZENGENTIK & KULTURPFLANZENFOR.
PI Bul M, Foerster S, Kunze G, Kunze I;
DR WPI; 96-065579/08.
PT Prodn. of heat stable glucoamylase in yeast able to utilise starch - by
PT transformation with glucoamylase gene including secretion sequence from
PT Arxula adeninivorans
PS Claim 1; Page 7; 21pp; German.
CC The present sequence is the N-terminal sequence of a heat stable
CC glucoamylase derived from Arxula adeninivorans. Yeast, e.g.
CC Saccharomyces cerevisiae, can be transformed with the DNA (T08701)
CC encoding the glucoamylase (R77674) for the production of the enzyme
CC which is able to used starch as its carbon source. The enzyme is useful
CC for degradation of starch, giving glucose as a final product. In
CC yeast the DNA encoding the enzyme can be easily manipulated to
CC produce the protein in high yield, which is otherwise difficult to
CC do in A.adeninivorans.
SQ Sequence 16 AA;

Query Match 42.0%; Score 55; DB 17; Length 16;
Best Local Similarity 50.0%; Pred. No. 1.81e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1 mrqflalaaas 12
   |||:| | : |
OY 5 MROYLVLMSQSS 16

```

Search completed: Fri Apr 16 14:21:38 1999
 Job time : 16 secs.

 Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

(TW)

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Apr 16 14:22:58 1999; Maspar time 5.06 Seconds
 Tabular output not generated. 167.360 Million cell updates/sec

Title: >US-09-011-797-3
 Description: (1-17) from US09011797.pep
 Perfect Score: 131
 Sequence: 1 FSEFMROYLVLSMOSQ 17

Scoring table: PAM 150
 Gap 15

Searched: 165420 segs, 49795644 residues
 Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: spltembl6
 1:sp:archaea 2:sp:bacteria 3:sp:fungi 4:sp:human
 5:sp:invertebrate 6:sp:mammal 7:sp:mhc 8:sp:organelle
 9:sp:phase 10:sp:plant 11:sp:rodent 12:sp:unclassified
 13:sp:vertebrate 14:sp:virus

Statistics: Mean 28.477; Variance 45.782; scale 0.622

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	131	100.0	176	6	062647	
2	68	51.9	210	2	007229	
3	68	51.9	310	2	051297	
4	65	49.6	193	10	064946	
5	65	49.6	339	10	040063	
6	64	48.9	310	10	004076	
7	64	48.9	312	1	058550	
8	63	48.1	169	10	040062	
9	62	47.3	91	5	061572	
10	62	47.3	320	10	024108	
11	62	47.3	406	2	066456	
12	61	46.6	118	2	054885	
13	61	46.6	312	11	097544	
14	61	46.6	317	10	064947	
15	61	46.6	733	4	049567	
16	61	46.6	792	4	060785	
17	60	45.8	320	10	040399	
18	60	45.8	434	2	005390	
19	60	45.8	435	2	032056	
20	60	45.8	435	2	032056	

21	60	45.8	700	11	054843	80DA M-CALPAIN SUBUNIT	2.14e+01
22	59	45.0	126	1	030234	HYPOTHETICAL 14.5 KD P	3.05e+01
23	59	45.0	261	3	007508	CHROMOSOME IV READING	3.05e+01
24	59	45.0	265	2	068268	WAP.	3.05e+01
25	59	45.0	317	10	041681	1-AMINOCYCLOPROPANE-1-	3.05e+01
26	59	45.0	321	10	039683	AMINO-CYCLOPROPANE CAR	3.05e+01
27	59	45.0	358	10	041698	CISTEINE PROTEINASE PR	3.05e+01
28	59	45.0	365	2	045135	HOMOPROTEOCHUATE 2,	3.05e+01
29	59	45.0	431	10	049566	HYPOTHETICAL 46.7 KD P	3.05e+01
30	59	45.0	923	2	053890	PGRS-FAMILY PROTEIN.	3.05e+01
31	59	45.0	963	10	039159	MYOSIN HEAVY CHAIN (FR	3.05e+01
32	59	45.0	3443	14	011979	POLYPROTEIN.	3.05e+01
33	58	44.3	95	14	05878	VIRULENCE DETERMINANT.	4.32e+01
34	58	44.3	95	14	055883	VIRULENCE DETERMINANT.	4.32e+01
35	58	44.3	95	14	055879	VIRULENCE DETERMINANT.	4.32e+01
36	58	44.3	95	14	055877	VIRULENCE DETERMINANT.	4.32e+01
37	58	44.3	299	10	049903	ACC OXIDASE.	4.32e+01
38	58	44.3	307	10	039933	1-AMINOCYCLOPROPANE-1-	4.32e+01
39	58	44.3	315	10	093326	1-AMINOCYCLOPROPANE-1-	4.32e+01
40	58	44.3	319	10	043792	ETHYLENE FORMING ENZY	4.32e+01
41	58	44.3	319	10	043581	ETHYLENE FORMING ENZY	4.32e+01
42	58	44.3	320	10	065378	FLZEL 12 PROTEIN.	4.32e+01
43	58	44.3	434	13	057391	ENOLASE (EC 4.2.1.11)	4.32e+01
44	58	44.3	692	2	094281	GYRB.	4.32e+01
45	58	44.3	1280	5	001779	COSMID W03F11.	4.32e+01

ALIGNMENTS

RESULT	ID	Score	Length	DB	Description	Pred. No.
1	062647	100.0%	176	6	062647	
2	01-AUG-1998 (TREMURREL. 07, CREATED)	100.0%	176	6	062647	
3	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
4	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
5	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
6	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
7	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
8	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
9	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
10	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
11	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
12	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
13	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
14	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
15	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
16	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
17	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
18	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
19	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
20	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
21	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
22	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
23	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
24	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
25	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
26	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
27	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
28	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
29	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
30	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
31	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
32	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
33	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
34	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
35	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
36	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
37	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
38	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
39	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
40	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
41	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
42	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
43	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
44	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	
45	01-AUG-1998 (TREMURREL. 07, LAST SEQUENCE UPDATE)	100.0%	176	6	062647	

Query Match 100.0%; Score 131; DB 6; Length 176;
 Best Local Similarity 100.0%; Pred. No. 4.71e-13;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2 PRELIMINARY; PRT; 210 AA.
 ID 007229;
 AC 007229;
 DT 01-JUL-1997 (TREMURREL. 04, CREATED)
 DT 01-JUL-1997 (TREMURREL. 04, LAST SEQUENCE UPDATE)
 DE PUTATIVE REGULATOR PROTEIN MTC63.07.
 GN MTC63.07.
 OS MYCOBACTERIUM TUBERCULOSIS.
 OC PROKARYOTA; FIRMITICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-H37RV.
 RA DEVLIN K., CHURCHER C.M.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALSUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL: 296800; E321650;
 DR PFAM: PF00440; tetr.
 SQ SEQUENCE 210 AA; 23396 MW; 288662A CRC32;
 Query Match 51.9%; Score 68; DB 2; Length 210;
 Best Local Similarity 43.8%; Pred. No. 1.10e+00;
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 Db 106 PDELRFLYLSMERGQ 121
 QY 2 SEFMROYLVLSMOSSQ 17

RESULT 3
 ID 051297. PRELIMINARY; PRT: 310 AA.
 AC 051297.
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
 GN B80317.
 OS BORELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;
 OC SPIROCHAETACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE: 98065943.
 RA FRASER C.M., CASSENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LAHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALBERG S., HANSON M.,
 RA VAN VUUT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTERBACK T., MATTHEY L., McDONALD L., ARTACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RL NATURE 390:580-586(1997).
 DR EMBL: AE00138; G2688213; -.
 DR TIGR: B80317; -.
 SQ SEQUENCE 310 AA; 35680 MW; C8F8F32F CRC32;

Query Match 51.9%; Score 68; DB 2; Length 310;
 Best Local Similarity 50.0%; Pred. No. 1.10e+00;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Db 228 FVFFRPYLVGLSG 243
 QY 1 FSEFMROYLVLSMOSS 16

RESULT 4
 ID 064946. PRELIMINARY; PRT: 193 AA.
 AC 064946;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ACC OXIDASE 2 (FRAGMENT).
 GN CS-ACO2.
 OS CUCUMIS SATIVUS (CUCUMBER).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC VIOLALES; CUCURBITACEAE.
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-FLORAL BUDS;
 RA RYAN J.E., ANDRALOJC P.J., WILLIS A.C., GUTTERIDGE S., PARRY M.A.J.;
 RL PLANT PHYSIOL. 116:1192-1192(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FLORAL BUDS;
 RA PERL-TREVES R., KAHANA A., KORACH T., KESSLER N.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF033582; G3025895; -.
 FT NON-TER
 SQ SEQUENCE 193 AA; 21553 MW; 1D8F81D1 CRC32;
 Query Match 49.6%; Score 65; DB 10; Length 193;
 Best Local Similarity 29.4%; Pred. No. 3.45e+00;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 155 FDDYMKLYMGKIFQAKE 171
 QY 1 FSEFMROYLVLSMOSSQ 17

RESULT 5
 ID 040063. PRELIMINARY; PRT: 339 AA.
 AC 040063;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE IDS3
 OS HORDEUM VULGARE (BARLEY).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 OC CYPERALES; GRAMINEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VAR. NK 1558;
 RA NAKANISHI H., SAKAGUCHI T., OKUMURA N., UMEHARA Y., NISHIZAWA N.,
 RA CHINO M., MORI S.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: D37796; G520582; -.
 DR PFAM: PF00671; Fe_Asc-oxidored.
 SQ SEQUENCE 339 AA; 37732 MW; B38285E1 CRC32;

Query Match 49.6%; Score 65; DB 10; Length 339;
 Best Local Similarity 50.0%; Pred. No. 3.45e+00;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Db 310 FRDEMRIVNVVLGSS 325
 QY 1 FSEFMROYLVLSMOSS 16

RESULT 6
 ID 004076. PRELIMINARY; PRT: 310 AA.
 AC 004076;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE ACC-OXIDASE.
 GN AB-ACOL.
 OS VIGNA ANGIUARS (ADZUKI BEAN).
 OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; VIRIDIPANTAE;
 OC CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA; MAGNOLIOPHYTA;
 OC MAGNOLIOPSIDA; RUTANAE; SAPINDALES; FABACEAE; PAPILIONOIDEAE;
 OC VIGNA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EPICOTYLS.
 RA KANEITA T., KAKIMOTO T., SHIBAKA H.;
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AB002667; D1020382; -.
 DR PFAM: PF00671; Fe_Asc-oxidored.
 SQ SEQUENCE 310 AA; 35334 MW; C9035224 CRC32;

Query Match	48.9%	Score 64	DB 10	Length 310
Best Local Similarity	35.3%	Pred. No.	5.01e+00	
Matches	6	Conservative	4	Mismatches 7
				Indels 0
				Gaps 0
Db	282	FEDYMKLYATLKFQPK	298	
		::	:	
	1	FSSEFQRYLVLSKSSQ	17	
		oy		

RESULT	7		
ID	058550	PRELIMINARY;	PRT; 312 AA.
AC	058550:		
DT	01-AUG-1998	(TREMBLREL. 07, CREATED)	
DT	01-AUG-1998	(TREMBLREL. 07, LAST SEQUENCE UPDATE)	
DT	01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)	
DE	312AA LONG HYPOTHERICAL PROTEIN.		
GN	PHC0033.		
OS	PYROCOCCLUS HORIKOSHII.		
OC	ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-073		
RA	KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,		
RA	YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,		
RA	SAKAI M., OGURA K., OTSUKA R., NAKAZAKA H., TAKAMITA M., OHEFUKU Y.,		
RA	FUJIMASHI T., TANAKA T., KUDOH Y., YAMAZAKI N., KUSHIDA N., OGUCHI A.,		
RA	AOKI K., YOSHIZAWA T., NAKAMURA Y., MASUCHI Y., SHIZUYA H.,		
RA	KIKUCHI H.;		
RL	SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBD DATA BANKS.		
SR	EMBL, AB009492; D1027933; --		
SO	SEQUENCE 312 AA; 35239 MW; A56CAE44 CRC32;		

RESULT	8	PRELIMINARY:	PRT:	169 AA.
ID	040062			
AC	040062:			
DT	01-NOV-1996	(TREMBLEL. 01, CREATED)		
DT	01-NOV-1996	(TREMBLEL. 01, LAST SEQUENCE UPDATE)		
DT	01-AUG-1998	(TREMBLEL. 07, LAST ANNOTATION UPDATE)		
DE	IRON-DEFICIENCY SPECIFIC.			
GN	IDS-3.			
OS	HORDEUM VULGARE (BARLEY).			
OC	EUKARYOTA: PLANTAE: EMBRYOPHYTA: ANGIOSPERMAE: MONOCOTYLEDONEAE:			
CC	CYPERALES: GRAMINEAE.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-ROOT:			
RA	MORI S.:			
RL	SUBMITTED (APR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.			
LN	(2)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-ROOT:			
RA	NAKANISHI H., OKUMURA N., UMEHARA Y., NISHITAZAWA N.K., CHINO M.,			
RL	PLANT CELL PHYSIOL. 34:401-410(1994).			
DR	EMBL: D10058; G217877; -			
SO	SEQUENCE 169 AA; 18704 MW; 83B15F7B CRC32;			

RESULT	ID	PRELIMINARY:	PRT:	91 AA.
	AC 061572;			
DT	01-AUG-1998 (TREMBLREL, 07, CREATED)			
DT	01-AUG-1998 (TREMBLREL, 07, LAST SEQUENCE UPDATE)			
DT	01-AUG-1998 (TREMBLREL, 07, LAST ANNOTATION UPDATE)			
DE	HYPOTHEITICAL 10.4 KD PROTEIN.			
OS	OSTERTAGIA OSTERTAGI.			
OC	EMBARCOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHADITIDA			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MOORE J., DEVANEY E.;			
RL	SUBMITTED (MAR-1998) TO EMBL/GENEBANK/DBJ DATA BANKS.			
DR	EMBL; AF052047; G2981080; -			
KW	HYPOTHEITICAL PROTEIN.			
SO	SEQUENCE 91 AA; 10442 MW; 6448C6C7 CRC32;			
	Query Match	47.3%;	Score 62;	DB 5; Length 91;
	Best Local Similarity	35.3%;	Pred. No. 1.04e+01;	
Matches	6; Conservative		7; Mismatches	4; Indels
			0; Gaps	
Db	44 YEERFRRYLVLEDSTE 60			
Y	1 FSEFRRIYLVLSMOSSO 17			

```

RESULT      10
ID           024108      PRELIMINARY;      PRT;      320 AA.
AC           024108:
DT           01-JAN-1998 (TREMBREL. 05, CREATED)
DT           01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DT           01-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)
DE           1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE.
GN           NGACO1.
OS           NICOTIANA GLUTINOSA (TOBACCO).
OC           EMBRYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEONEAE
OC           SOLANALES; SOLANACEAE.
RN           [1]
RP           SEQUENCE FROM N.A.
RC           TISSUE-LEAF;
RA           LEE S.H., KIM Y.S., LEE M.M., CHOI D.I., KIM W.T.;
RL           SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR           EMBL; U54565; G1314707;
PF           PF00671; Fe4SC4O4 oxidored.
SQ           SEQUENCE 320 AA; 36288 MW;  FDDDEB82 CRC32;

```

RESULT	11	
ID	066456	PRELIMINARY; PRT. 406 AA.
AC	066456;	
DT	01-AUG-1998	(TREMBLREL. 07, CREATED)
DT	01-AUG-1998	(TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT	01-AUG-1998	(TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE	HYPOTHETICAL 47.6 KD PROTEIN.	
GN	AQ.040.	
OS	AQUIFEX AEOLICUS.	
OC	EUBACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-VF5;	
RA	DECKERT G., WARREN P. V., GAASTERLAND T., YOUNG W. G., LENOX A. L.,	
RA	GRAHAM D. E., OVERBEER K., SNEAD M. A., KELLER M., ADJAY M., HUBER	
RA	FELDMAN R. A., SHORT J. M., OLSON G. J., SWANSON R. V.;	

RL NATURE 392:353-358(1998).
 (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-VES;
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AUYAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: A600670; G2982788; -
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 406 AA; 47559 MW; 3A69FCF8 CRC32;
 Query Match 47.3%; Score 62; DB 2; Length 406;
 Best Local Similarity 63.6%; Pred. No. 1.04e+01;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Db 191 FEEFRETIVL 201
 QY 1 FSEFMROYLV 11
 RESULT 12
 ID 054885 PRELIMINARY; PRT; 118 AA.
 AC 054885;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ORF1.
 OS STREPTOCOCCUS PNEUMONIAE.
 SC PROKARYOTA; BACTERIA; FIRMICUTES; GRAM-POSITIVE COCCI;
 OC STREPTOCOCCACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSOM-TRANSPOS TNS252;
 RA VIJAYAKUMAR M.N.;
 RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: L29324; G951053; -
 SQ SEQUENCE 118 AA; 14251 MW; 1C07C914 CRC32;
 Query Match 46.6%; Score 61; DB 2; Length 118;
 Best Local Similarity 58.3%; Pred. No. 1.50e+01;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 34 FSDFLRKLKLLS 45
 QY 1 FSEFMROYLVLS 12
 RESULT 13
 ID P97544 PRELIMINARY; PRT; 312 AA.
 AC P97544;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE ER TRANSMEMBRANE PROTEIN.
 GN DRI 42.
 OS RATTUS NORVEGICUS (RAT).
 SC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MISTAR; TISSUE=SMALL INTESTINE;
 RX MEDLINE: 97094703.
 RA BARIJA D., PLATEROTI M., NOBILI F., MUDA A.O., XIE Y., MORIMOTO T.,
 RA PEROZZI G.;
 RL J. BIOL. CHEM. 271:29928-29936(1996).
 DR EMBL: Y07783; E283078; -
 KW TRANSMEMBRANE.
 SQ SEQUENCE 312 AA; 35318 MW; 5A8F58E2 CRC32;
 Query Match 46.6%; Score 61; DB 11; Length 312;
 Best Local Similarity 53.3%; Pred. No. 1.50e+01;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 203 FSNFTMLVLYLQA 217
 QY 1 FSEFMROYLVLSMOS 15
 RESULT 14
 ID 064947 PRELIMINARY; PRT; 317 AA.
 AC 064947;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ACC OXIDASE 3.
 GN CS-AC03.
 OS CUCUMIS SATIVUS (CUCUMBER).
 SC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC VIOLALES; CUCURBITACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FLORAL BUDS;
 RA RYAN J.E., ANDRALOJC P.J., WILLIS A.C., GUTTERIDGE S., PARRY M.A.J.;
 RL PLANT PHYSIOL. 116:1192-1192(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FLORAL BUDS;
 RA PERL-TREVES R., KAHANA A., KORACH T., KESSLER N.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF033583; G3025697; -
 SQ SEQUENCE 317 AA; 35875 MW; 83DC530F CRC32;
 Query Match 46.6%; Score 61; DB 10; Length 317;
 Best Local Similarity 29.4%; Pred. No. 1.50e+01;
 Matches 5; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 Db 283 FEDYMKLYGVKFOAKE 299
 QY 1 FSEFMROYLVLSMOSQ 17
 RESULT 15
 ID 049567 PRELIMINARY; PRT; 393 AA.
 AC 049567;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 43.5 KD PROTEIN.
 GN F7J7.200.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 SC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC CAPRARIAS; CROCIIFERAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., MURPHY G., DROST L., HALL C., HUDSON S., RIDLEY P.,
 RA BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL EMBL: AL021960; E1253322; -
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 393 AA; 43457 MW; 3806ED3E CRC32;
 Query Match 46.6%; Score 61; DB 10; Length 393;
 Best Local Similarity 54.5%; Pred. No. 1.50e+01;
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 Db 311 OYLIGMONT 321
 QY 7 QYLVLSMOSQ 17
 Search completed: Fri Apr 16 14:23:33 1999

Wed Apr 28 09:08:36 1999

US-09-011-797-3.rspt

Page 5

Job time : 35 secs.

200

THIS PAGE BLANK (USPTO)